Ring Arithmetic, Field Extensions
and Applications
Table of Contents

Foreword.................................................................................................................................7

I. Arithmetic in integral domains ..........................................................................................10
   I.1 Divisibility..................................................................................................................10
   I.2 Euclidian domains ......................................................................................................22
   I.3 Euclidian rings of quadratic integers ........................................................................26
   I.4 Principal ideal domains ..............................................................................................31
   I.5 Unique factorization domains.....................................................................................36
   I.6 Polynomial ring arithmetic.........................................................................................43
   Exercises.............................................................................................................................50

II. Modules................................................................................................................................58
   II.1 Modules, submodules, homomorphisms.................................................................58
   Exercises............................................................................................................................72
   II.2 Factor modules and the isomorphism theorems....................................................74
   Exercises............................................................................................................................80
   II.3 Direct sums and products. Exact sequences .........................................................81
   Exercises............................................................................................................................102
   II.4 Free modules.............................................................................................................104
   Exercises............................................................................................................................116

III. Finitely generated modules over principal ideal domains..................118
   III.1 The submodules of a free module...........................................................................118
   Exercises............................................................................................................................128
   III.2 Finitely generated modules over a principal ideal domain................................130
   Exercises............................................................................................................................138
   III.3 Indecomposable finitely generated modules .......................................................140
   Exercises............................................................................................................................149
III.4 The endomorphisms of a finite dimensional vector space ..... 150
Exercises ........................................................................................................ 167

IV. Field extensions ...................................................................................... 169
IV.1. Algebraic extensions ........................................................................ 169
Exercises ........................................................................................................ 190
IV.2 Roots of polynomials. Algebraically closed fields ............................ 193
Exercises ........................................................................................................ 207
IV.3 Finite fields ......................................................................................... 209
Exercises ........................................................................................................ 221
IV.4 Transcendental extensions .................................................................. 223
Exercises ........................................................................................................ 230

V. Galois Theory ......................................................................................... 232
V.1 Automorphisms ................................................................................... 233
V.2 Normal extensions ................................................................................ 238
Exercises ........................................................................................................ 242
V.3 Separability .......................................................................................... 244
Exercises ........................................................................................................ 259
V.4 The Fundamental Theorem of Galois Theory .................................. 262
Exercises ........................................................................................................ 271

VI. Applications of Galois Theory ............................................................ 273
VI.1 Ruler and compass constructions .................................................. 273
VI.2 Trace and norm .................................................................................. 289
Exercises ........................................................................................................ 296
VI.3 Cyclic extensions and Kummer extensions ...................................... 297
Exercises ........................................................................................................ 312
VI.4 Solvability by radicals ....................................................................... 313
VI.5 Discriminants, resultants .................................................................... 320
Exercises ........................................................................................................ 326

Appendices .................................................................................................. 331
1. Prime ideals and maximal ideals ......................................................... 331
2. Algebras. Polynomial and monoid algebras ...................................... 333
3. Symmetric polynomials ................................................................. 348
4. Rings and modules of fractions ......................................................... 354
5. Categories, functors ............................................................................. 362
6. Solvable groups .................................................................................... 373

Index ............................................................................................................ 379
Bibliography ............................................................................................... 387
Foreword

The book is aimed at undergraduate students in Mathematics, Computer Science or technical universities having a background of standard courses of Abstract Algebra and Linear Algebra as well as at general Mathematics readers with an interest in Algebra.

The topics covered by the book are mandatory algebraic background of any mathematics graduate: arithmetic in integral domains, module theory basics, the structure of the finitely generated modules over a principal ideal domain with applications in abelian groups and Jordan forms of matrices, field extensions and Galois theory with applications.

Throughout the book the reader is motivated by concrete applications and exercises.

The book is reasonably self contained, in the sense that the reader is assumed to be familiar with general notions on algebraic structures (monoids, groups, rings, fields), factor rings and isomorphism theorems, vector spaces and bases, matrices, polynomials, permutation groups basics, elementary arithmetic of cardinals. Some topics less likely to appear in standard general Algebra courses are presented in the Appendix.

The language of Category Theory is used beginning with the chapter on Modules, the reader being invited to refer to the appendix (or to a book on Categories) when categorical notions appear in the text. We
think that the Categories are particularly useful for a better understanding and for unifying many algebraic concepts and proofs.

The chapters on Modules (II, III) can be read independently of the rest of the book. The section VI.1, Ruler and compass constructions, is not a prerequisite for the other sections in chapter VI.

Some notations used in the text:
- $|A|$ denotes the cardinal of the set $A$ (the number of elements of $A$, if $A$ is finite).
- $x := y$ means “$x$ is equal by definition to $y$” (where $y$ is already defined) or “we denote $y$ with $x$"
- $\square$ marks the end or the absence of a proof.
- $\mathbb{N}$ is the set of natural numbers, \{0, 1, 2, \ldots\}
- $\mathbb{N}^*$ is the set of positive natural numbers, \{1, 2, \ldots\}
- $\mathbb{Z}$ is the set of integers
I. Arithmetic in integral domains

The set \( \mathbb{Z} \) of integers, endowed with the operations of addition and multiplication, is the prototype for the familiar concept of \textit{ring}. The classical divisibility theory in \( \mathbb{Z} \) (the \textit{arithmetic} of \( \mathbb{Z} \)) can be extended with outstanding results to a large class of rings, the \textit{integral domains}. Such a generalization is interesting by itself and it also illuminates and yields nontrivial results on the divisibility in \( \mathbb{Z} \).

After a general study of the divisibility in integral domains, three classical and important classes of domains are studied (\textit{Euclidian domains}, \textit{principal ideal domains} and \textit{unique factorization domains}). The definitions of these classes of rings originate in fundamental arithmetic properties of \( \mathbb{Z} \). The material in this chapter is at the very basis of all Algebra, and vital in Algebraic Number Theory, Field Extensions and Galois Theory.

I.1 Divisibility

The classical definition for the relation of divisibility in the ring of integers \( \mathbb{Z} \) generalizes easily to an arbitrary ring \( R \):
1.1 Definition. Let $R$ be a ring and let $a, b \in R$. We say that $a$ divides $b$ in $R$ (and write $a \mid b$) if there exists $c \in R$ such that $b = ac$.

The fact that $a \mid b$ can be also expressed by writing $b : a$ (read “$b$ is divisible by $a$”). Other ways of reading $a \mid b$ are: “$a$ is a divisor of $b$” or “$b$ is a multiple of $a$”.

One says “$a$ divides $b$ in $R$” because the ring $R$ plays an essential role here. For instance, $2 \mid 3$ in $\mathbb{Q}$, but of course not in $\mathbb{Z}$! We shall omit any reference to $R$ in the notation $a \mid b$ if the ring $R$ is clear from the context. Write $a \nmid b$ if $a$ does not divide $b$.

If the ring $R$ lacks some natural properties, the theory of divisibility in $R$ can be very poor or very peculiar when compared to the classical theory in $\mathbb{Z}$. For instance, in a ring without identity element, an element may not divide itself; other difficulties arise if $R$ is not commutative or if $R$ has zero divisors.

This motivates the following definition:

1.2 Definition. A ring $R$ is called an integral domain (a domain, for short) if it has identity (denoted by 1), it is commutative and has no zero divisors: for any $x, y \in R$, $xy = 0$ implies $x = 0$ or $y = 0$. This can also be said: for any nonzero $x, y \in R$, we have $xy \neq 0$.

In what follows, all rings we consider are domains, 0 denotes the zero element of the domain and 1 its identity element. All subrings considered are unitary (they contain the identity element of the ring).

1.3 Examples. a) Any field $F$ (for instance $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, ...) is a domain. Indeed, if $x, y \in F$ are nonzero, then $xy = 0$ implies (by multiplying with $x^{-1}$, which exists in $F$ since $F$ is a field and $x \neq 0$), that $y = 0$, contradiction. The theory of divisibility in fields is trivial, though (cf. 1.7).

b) Every subring of a domain is itself a domain. In particular, every subring of a field is a domain. So, if $d \in \mathbb{Z}$ is squarefree (i.e.: $d \neq 0$,
I. Arithmetic in integral domains

\[ d \neq 1 \text{ and } d \text{ is not divisible by the square of any integer greater than } 1 \), the subring of \( \mathbb{C} \) generated by 1 and \( \sqrt{d} \), denoted by \( \mathbb{Z} [\sqrt{d}] \), is a domain. One easily checks that \( \mathbb{Z} [\sqrt{d}] \) consists of the complex numbers of the form \( a + b\sqrt{d} \), with \( a, b \in \mathbb{Z} \). The ring \( \mathbb{Z} [\sqrt{-1}] \) is called the ring of Gauss integers.\(^1\)

c) If \( R \) is a domain and \( n \in \mathbb{N}^* \), then the polynomial ring in \( n \) indeterminates with coefficients in \( R \), \( R[X_1, \ldots, X_n] \), is a domain.

In a domain, one can simplify the nonzero factors:

**1.4 Proposition.** Let \( R \) be a domain and let \( a, b, c \in R \), with \( c \neq 0 \). If \( ac = bc \), then \( a = b \).

**Proof.** We have \( ac = bc \iff ac - bc = 0 \iff (a - b)c = 0 \). Since \( R \) is a domain, \( a - b = 0 \) or \( c = 0 \). But \( c \neq 0 \), so \( a - b = 0 \).

We will develop a theory of divisibility in \( R \), with \( \mathbb{Z} \) as a model (and a particular case). The proof of the following properties is an easy exercise:

**1.5 Proposition.** Let \( R \) be a domain. Then:

a) For any \( a \in R \), \( a \mid a \).

b) For any \( a, b, c \in R \) such that \( a \mid b \) and \( b \mid c \), we have \( a \mid c \).

c) For any \( a \in R \), we have \( a \mid 0 \) and \( 1 \mid a \).

d) For any \( x, y \in R \) and \( a, b, c \in R \) such that \( a \mid b \) and \( a \mid c \), we have \( a \mid (bx + cy) \).

Properties a) and b) say that the divisibility relation on \( R \) is reflexive and transitive.

**1.6 Definition.** The elements \( a, b \) in \( R \) are called associated in divisibility (or, simply, associated) if \( a \mid b \) and \( b \mid a \). Notation: \( a \sim b \).

\(^1\) Carl Friedrich Gauss (1777 – 1855), famous German mathematician.
For \( d, a \in R \), \( d \) is called a proper divisor of \( a \) if \( d \mid a \) and \( d \) is neither invertible in \( R \), nor associated with \( a \).

The relation "\( \sim \)" defined above is an equivalence relation on \( R \) (exercise!) and it is very important when studying the arithmetic of \( R \): two elements associated in divisibility have exactly the same divisors and the same multiples. One can say they are indistinguishable as far as divisibility is concerned.

An invertible element \( u \in R \) is called a unit of \( R \), because \( u \sim 1 \) (so \( u \) behaves just like 1 from the divisibility standpoint). Let \( U(R) \) denote the set of all invertible elements of \( R \):

\[
U(R) = \{ x \in R \mid (\exists) \ y \in R \text{ such that } xy = 1 \}\.
\]

\( U(R) \) is a group with respect to the ring multiplication (as a straightforward checking shows) and is called the group of units of \( R \).

**1.7 Proposition.** Let \( R \) be a domain. Then :

a) For any \( u \in R \), we have: \( u \in U(R) \iff u \sim 1 \iff u \mid a, \ (\forall) \ a \in R \iff uR = R \).

b) For any \( a, b \in R \), we have: \( a \sim b \iff \text{there exists } u \in R \text{ such that } a = bu. \)

For a given domain \( R \), knowing the group \( U(R) \) is a first step, very important, in the study of divisibility in \( R \).

**1.8 Examples.**

a) \( U(\mathbb{Z}) = \{-1, 1\} \).

b) If \( K \) is a field, \( U(K[X]) = \{ f \in K[X] \mid \deg f = 0 \} = K^\ast \) (we identify nonzero elements in \( K \) with the polynomials of degree 0).

c) If \( d \in \mathbb{Z} \) is squarefree, then:

\[
U(\mathbb{Z}[\sqrt{d}]) = \{ a + b\sqrt{d} \mid a, b \in \mathbb{Z}, a^2 - db^2 = \pm 1 \}
\]

**Proof.** Let \( R = \mathbb{Z}[\sqrt{d}] \). It is useful to define the “norm” \( N : R \rightarrow \mathbb{Z} \), by \( N(\alpha) = \alpha \sigma(\alpha) \), where \( \sigma(\alpha) \) is the conjugate of \( \alpha \), defined as:

\[
\sigma(a + b\sqrt{d}) = a - b\sqrt{d} , \text{ for any } a, b \in \mathbb{Z}.
\]
I. Arithmetic in integral domains

So \( N(a + b\sqrt{d}) = a^2 - db^2 \), \( \forall a, b \in \mathbb{Z} \). An easy computation shows:

\[ N(\alpha)N(\beta) = N(\alpha\beta), \forall \alpha, \beta \in R. \]

This implies: if \( \alpha, \beta \in R \) with \( \alpha \mid \beta \) in \( R \), then \( N(\alpha) \mid N(\beta) \) in \( \mathbb{Z} \).

Let \( u = a + b\sqrt{d} \in U(R) \). Then \( N(u) = a^2 - db^2 \) divides 1 in \( \mathbb{Z} \), so \( N(u) = \pm 1 \). Conversely, if \( N(u) = \pm 1 \), then \( (a + b\sqrt{d})(a - b\sqrt{d}) = \pm 1 \), so \( \pm \left( a - b\sqrt{d} \right) \) is the inverse of \( u \). Thus:

\[
U(\mathbb{Z}[\sqrt{d}]) = \{ a + b\sqrt{d} \mid a, b \in \mathbb{Z}, a^2 - db^2 = \pm 1 \} = \\
\{ \alpha \in \mathbb{Z}[\sqrt{d}] \mid N(\alpha) = \pm 1 \}.
\]

We define the central concepts of greatest common divisor and least common multiple. Since an order relation like the one on \( \mathbb{N} \) is not available, we use the divisibility relation itself to order the common divisors.

**1.9 Definition.** Assume \( R \) is a domain, \( n \in \mathbb{N}^* \) and \( a_1, \ldots, a_n \in R \). We call the element \( d \in R \) a greatest common divisor (abbreviated GCD) of the elements \( a_1, \ldots, a_n \) if:

i) \( d \mid a_1, \ldots, d \mid a_n \). \( (d \) is a common divisor of \( a_1, \ldots, a_n \)

ii) For any \( e \in R \) such that \( e \mid a_1, \ldots, e \mid a_n \), it follows that \( e \mid d \) \( (d \) is the “greatest” among the common divisors of \( a_1, \ldots, a_n \).

If 1 is a GCD of \( a_1 \) and \( a_2 \), we call \( a_1 \) and \( a_2 \) coprime or mutually prime (or relatively prime).

We call \( m \in R \) a least common multiple (LCM for short) of \( a_1, \ldots, a_n \) if:

i') \( a_1 \mid m, \ldots, a_n \mid m \).

ii') For any \( e \in R \) such that \( a_1 \mid e, \ldots, a_n \mid e \), it follows that \( m \mid e \).

We denote by \( (a_1, \ldots, a_n) \) or \( \text{GCD}(a_1, \ldots, a_n) \) a GCD of \( a_1, \ldots, a_n \), if it exists.

Similarly, \([a_1, \ldots, a_n] \) or \( \text{LCM}(a_1, \ldots, a_n) \) denotes a LCM of \( a_1, \ldots, a_n \), if it exists.
1.10 Remarks. a) Given \( a_1, \ldots, a_n \in R \), if there exists a GCD for \( a_1, \ldots, a_n \), say \( d \in R \), then \( d \) is uniquely determined up to association in divisibility: if \( e \) is also a GCD of \( a_1, \ldots, a_n \), then \( e \sim d \). Moreover, if \( e \sim d \), then \( e \) is a GCD of \( a_1, \ldots, a_n \).

The same remark applies to the LCM.

b) When the domain \( R \) in which we work is not clear from the context, we use occasionally a subscript, like in the notation \((a_1, \ldots, a_n)_R\).

For a given domain \( R \) and given \( x, y \in R \), a GCD \((x, y)\) may not exist (see the Exercises for some examples). A domain \( R \) with the property that any two elements \( x, y \in R \) possess a GCD is called a GCD domain. For instance, \( \mathbb{Z} \) is a GCD domain.

Writing \( d = (a_1, \ldots, a_n) \) means that \( d \) is associated with a GCD of \( a_1, \ldots, a_n \). This can lead to some oddities: in \( \mathbb{Z} \), we can write \( 1 = (1, 2) = -1 \), but this does not imply that \( 1 = -1 \) (of course, it implies that \( 1 \sim -1 \)).

c) Note that \( a_1, a_2 \) are coprime if and only if all their common divisors are units in \( R \).

d) For any domain \( R \) and any \( a \in R \), there exists GCD \((a, 0) = a\). If \( u \) is a unit, then there exists GCD \((a, u) = u\). What can you say about the LCM in these cases?

If \( d \neq 0 \) and \( d|a \), \( a/d \) denotes the unique element \( x \in R \) with \( a = dx \).

1.11 Proposition. Let \( R \) be a domain and let \( a_1, \ldots, a_n, r \in R \setminus \{0\} \).

a) If there exists \( d = (a_1, \ldots, a_n) \), then \( a_1/d, \ldots, a_n/d \) have a GCD, equal to 1.

b) If there exists \((a_1, \ldots, a_n) =: d \) and exists \((ra_1, \ldots, ra_n) =: e \), then \( e = rd \). Thus:

\[
(ra_1, \ldots, ra_n) = r(a_1, \ldots, a_n).
\]

c) If there exists \([a_1, \ldots, a_n] = m \) and exists \([ra_1, \ldots, ra_n] =: \mu \), then \( \mu = rm \). Thus:

\[
[ra_1, \ldots, ra_n] = r[a_1, \ldots, a_n].
\]
I. Arithmetic in integral domains

Proof. a) Let \( x_i \in R \) such that \( a_i = dx_i, \ i = 1, \ldots, n \). If \( e \in R \) is a common divisor of \( x_1, \ldots, x_n \), then \( de \) is a common divisor of \( a_1, \ldots, a_n \), so \( de \mid d \). This implies \( e \mid 1 \).

b) Since \( rd \mid ra_i, \ i = 1, \ldots, n \), we get \( rd \mid e \). Let \( u \in R \) with \( e = rdu \). It is enough to prove that \( u \mid 1 \). Let \( x_i, y_i \in R \) such that \( a_i = dx_i \) and \( ra_i = ey_i, \ i = 1, \ldots, n \). For any \( i = 1, \ldots, n \), \( ra_i = rdx_i = rduy_i \). It follows that \( u \) is a common divisor of the elements \( x_i \), whose GCD is 1, by a). So, \( u \mid 1 \).

c) Because \( rm \) is a common multiple of \( ra_i, \ i = 1, \ldots, n \), we have \( \mu \mid rm \), so \( rm = \mu t \), for some \( t \in R \). We can write \( m = a_ib_i, \ \mu = rax_i, \) for some \( x_i, b_i \in R, \ i = 1, \ldots, n \). We have \( rm = ra_ib_i = \mu t = rax_it, \ i = 1, \ldots, n \). Simplifying, \( b_i = x_it \). Also, \( a_1x_1 = \ldots = a_nx_n \) is a common multiple of \( a_1, \ldots, a_n \). Thus, \( m \mid a_1x_i, \ i = 1, \ldots, n \). Since \( \mu = rax_i \), we also get \( mr \mid \mu \). □

1.12 Corollary. Let \( R \) be a GCD-domain and let \( K \) be the field of quotients of \( R \). Then every element in \( K \) can be written as a quotient \( \frac{a}{b} \), with \( a, b \in R, \ b \neq 0 \) and \( (a, b) = 1 \).

Proof. If \( \frac{c}{d} \in K \), with \( c, d \in R, \ d \neq 0 \), then let \( e = \text{GCD}(c, d) \). Then \( c = ea, \ d = eb \), for some \( a, b \in R \), with \( (a, b) = 1 \) (use the statement a) above). Moreover, \( \frac{c}{d} = \frac{a}{b} \). □

The next result is ubiquitous in divisibility arguments.

1.13 Corollary. Let \( R \) be a GCD domain and \( a, b, c \in R \) such that \( a \mid bc \) and \( (a, b) = 1 \). Then \( a \mid c \).

Proof. \( (a, b) = 1 \) and the preceding result, part b), imply that \( (ac, bc) = c \). Since \( a \mid ac \) and \( a \mid bc \), the definition of the GCD ensures that \( a \mid (ac, bc) = c \). □

Although the definitions of the GCD and the LCM are “dual” to each other, the situation is not entirely symmetric (the existence of the
LCM implies the existence of the GCD, but not conversely, in general).

1.14 Proposition. Let \( R \) be a domain and \( x, y \in R \). Then:

a) If a LCM of \( x, y \) exists, \([x, y] = m \in R\), then a GCD of \( x, y \) exists, \((x, y)\), and

\[ xy \sim [x, y](x, y) \]

b) If any two elements in \( R \) have a GCD, then any two elements in \( R \) have a LCM.

c) If any two elements in \( R \) have a GCD, then, for any \( n \in \mathbb{N}, n > 1 \), any \( n \) elements \( a_1, ..., a_n \) in \( R \) have a GCD and a LCM.

Proof. a) When \( x = 0 \), \([0, y] \) exists and it is 0. Similarly, \((0, y) = y\). Suppose now that \( x \) and \( y \) are nonzero. The definition of the LCM implies \( m \mid xy \). Let \( d, a, b \in R \) with \( xy = md \) and \( m = xa, m = yb \). We need only prove that \( d = (x, y) \). We have \( xy = xad \), so \( y = ad \Rightarrow d \mid y \). Likewise, \( d \mid x \). Take \( e \in R \) with \( e \mid x, e \mid y \) and pick \( r, s \in R \) such that \( x = er \) and \( y = es \). Then \( ers \) is a common multiple of \( x \) and \( y \), so \( m \mid ers \). Let \( t \in R \) such that \( mt = ers \). We have \( dm = xy = e^2rs = tem \). Simplifying by \( m \), \( d = te \), so \( e \mid d \).

b) Let \( a, b \in R \setminus \{0\} \) and let \( d = (a, b) \). There exist \( x, y \in R \) with \( a = dx, b = dy \). The element \( m = dxy \) is obviously a common multiple of \( a \) and \( b \). Let \( \mu \) be another common multiple of \( a \) and \( b \). There exist \( z, t \in R \) such that \( \mu = az = dxz \) and \( \mu = bt = dyt \). So \( m \) divides \( \mu y = dxyz \) and \( \mu x = dxyt \), which means \( m \) divides also \((\mu x, \mu y) = \mu(x, y) = \mu \). This shows that \( m \) is a LCM of \( a \) and \( b \).

c) Induction on \( n \). (Exercise!).

1.15 Example. In \( R = \mathbb{Z}[\sqrt{-5}] \), \( x = 1 + \sqrt{-5} \) and \( y = 2 \) have a GCD, but no LCM. Indeed, let \( d = a + b\sqrt{-5} \) \((a, b \in \mathbb{Z})\) be a common divisor of \( x \) and \( y \). Using the properties of the norm \( N \) (see 1.8.c.), we get \( N(d) \mid N(x) = 6 \) and \( N(d) \mid N(y) = 4 \) in \( \mathbb{Z} \). So, \( N(d) \mid 2 \) in \( \mathbb{Z} \). Since
\(N(d) = a^2 + 5b^2\), a case-by-case inspection leads to the conclusion that \(a = \pm 1\) and \(b = 0\). Thus, \(d\) is invertible. We proved that any common divisor of \(x\) and \(y\) is a unit, so \(x\) and \(y\) have GCD equal to 1.

Suppose a LCM \(\mu \in R\) of \(x\) and \(y\) exists. Then \(6 \mid N(\mu)\) and \(4 \mid N(\mu)\) in \(\mathbb{Z}\), so \(12 \mid N(\mu)\) in \(\mathbb{Z}\). On the other hand, \(6 = 2 \cdot 3 = (1 - \sqrt{5})(1 + \sqrt{5})\) and \(2(1 + \sqrt{5})\) are common multiples of \(x\) and \(y\), so they are common multiples of \(\mu\). Thus, \(N(\mu)\) divides \(N(6) = 36\) and \(N(2)N(1 + \sqrt{5}) = 24\) in \(\mathbb{Z}\), so \(N(\mu) \mid 12\). Combining with \(12 \mid N(\mu)\), we get \(N(\mu) = 12\), which is impossible (the equation \(a^2 + 5b^2 = 12\) has no solutions in \(\mathbb{Z}\)).

If \(R\) is a domain, let \(R^\circ\) designate the set of nonzero and non-invertible elements in \(R\):

\[
R^\circ := R \setminus \{0\} \setminus U(R)
\]

In \(\mathbb{Z}\), prime numbers play a central role in divisibility questions. Usually, the (elementary) definition for the notion of prime number is “the natural number \(p > 1\) is prime if its only divisors in \(\mathbb{N}\) are 1 and \(p\)”\). The generalization to the case of a domain of this definition leads to the notion of irreducible element (also compare with the notion of prime element below).

**1.16 Definition.** Let \(R\) be a domain.

The element \(p \in R^\circ\) is called irreducible (in \(R\)) if it has no proper divisors. In other words, any divisor of \(p\) is either a unit or is associated to \(p\): \(\forall d \in R, d \mid p \Rightarrow d \sim 1\) or \(d \sim p\).

The element \(p \in R^\circ\) is called prime (in \(R\)) if, for any \(a, b \in R, p \mid ab \Rightarrow p \mid a\) or \(p \mid b\).

We emphasize that a prime element or an irreducible element is by definition nonzero and non-invertible.
A quick argument shows that, for any $m \in \mathbb{N}^*$, if $p$ is prime and $p$ divides a product of $m$ factors in $R$, then $p$ divides one of the factors.

1.17 Proposition. Every prime element is also irreducible.

Proof. Let $p \in R$ be prime. If $d \in R$ is a divisor of $p$, there exists $x \in R$ (nonzero) such that $p = dx$. So $p \mid dx$, which implies $p \mid d$ (and we are finished) or $p \mid x$. But $p \mid x$ means that $p \sim x$ (since $x \mid p$), so $p = ux$, with $u$ a unit. So, $ux = dx = p$ and thus $u = d$ is a unit.

The notions of prime element and irreducible element (which coincide for $\mathbb{Z}$, as we will see) are not the same in general.

1.18 Example. In $\mathbb{Z}[\sqrt{-5}]$, 2 is irreducible and it is not prime. Indeed, 2 divides $(1 - \sqrt{-5})(1 + \sqrt{-5}) = 6$, but 2 divides neither factor. On the other hand, if $d$ is a divisor of 2, then $\text{N}(d)$ can only be 1, 2 or 4. An examination of the possible cases shows that $d$ is $\pm 1$ or $\pm 2$.

Thus, the notion of prime element depends heavily on the ring in which it is considered: 2 is prime in $\mathbb{Z}$, but not in $\mathbb{Z}[\sqrt{-5}]$. The same remark applies to the notion of irreducible element.

The GCD domains do not have the peculiarity described in the example above:

1.19 Proposition. Let $R$ be GCD domain. Then any irreducible element in $R$ is prime in $R$.

Proof. Let $p \in R$, irreducible and $x, y \in R$ such that $p \mid xy$. If $p \nmid x$, then the GCD of $p$ and $x$ (which exists!) is 1. Indeed, if $d \mid x$ and $d \mid p$, we cannot have $d \sim p$ (we would get $p \mid x$), so $d \sim 1$. Thus, $p \mid xy$ and $(p, x) = 1$. Corollary 1.13 guarantees that $p \mid y$.

The notion of divisibility can be translated in the language of ideals. This approach allows extending classical results on the divisibility
in \( \mathbb{Z} \) to much more general classes of rings (for instance, the primary decomposition theory).

Recall that a subset \( I \) of the commutative ring \( R \) is called an *ideal* of \( R \) if:

a) \((I, +)\) is a subgroup in the additive group \((R, +)\): \( \forall x, y \in I \Rightarrow x + y \in I \).

b) \( \forall x \in I, \forall r \in R \Rightarrow rx \in I \).

Write \( I \triangleleft R \) if \( I \) is an ideal in the ring \( R \). The ideal \( I \) is *proper* if \( I \neq R \).

For \( a \in R \), the ideal generated by \( a \) is the set \( \{ra \mid r \in R\} \), denoted by \( Ra \) or \( aR \) and is called the principal ideal generated by \( a \). The *sum* of two ideals \( I \) and \( J \) of \( R \) is the ideal:

\[
I + J := \{i + j \mid i \in I, j \in J\}.
\]

1.20 Proposition. Let \( R \) be a domain, \( n \in \mathbb{N}^* \) and \( a, b, x_1, \ldots, x_n \in R \). Then:

a) \( a \mid b \) if and only if \( Ra \supseteq Rb \).

b) \( a \sim b \) if and only if \( Ra = Rb \).

c) \( a \in U(R) \) if and only if \( Ra = R \).

d) \( a \) is prime in \( R \) if and only if \( Ra \) is a prime ideal.

e) \( a \) is irreducible in \( R \) if and only if \( Ra \) is a maximal ideal among the principal proper ideals of \( R \) (more precisely: \( \forall x \in R \) such that \( Ra \subseteq Rx \), we have \( Ra = Rx \) or \( Rx = R \)).

f) \( a \) is a common divisor of \( x_1, \ldots, x_n \) if and only if \( Rx_1 + \ldots + Rx_n \) is included in \( Ra \).

g) If \( Rx_1 + \ldots + Rx_n = Ra \), then \( a = (x_1, \ldots, x_n) \).\(^2\)

h) \( a \) is a common multiple of \( x_1, \ldots, x_n \) if and only if \( Rx_1 \cap \ldots \cap Rx_n \) includes \( Ra \).

\(^2\) The converse is false in general. For a counterexample, see the section Principal Ideal domains.
I.1 Divisibility

i) \(a = [x_1, \ldots, x_n]\) if and only if \(Rx_1 \cap \ldots \cap Rx_n = Ra\).

**Proof.**

a) \(a \mid b \iff \exists c \in R \text{ with } b = ca \iff b \in Ra \iff Rb \subseteq Ra\).

b) Obvious, by a).

c) If \(a\) is invertible, then \(\exists c \in R \text{ with } ca = 1\). So \(1 \in Ra \Rightarrow Ra = R\). Conversely, if \(Ra = R\), then \(1 \in Ra\), so there exists \(c \in R\) such that \(1 = ca\).

d) Let \(x, y \in R\). We have \(xy \in Ra \iff a \mid xy\). If \(a\) is prime, then \(a \mid x\) or \(a \mid y\), i.e. \(x \in Ra\) or \(y \in Ra\), which shows that \(Ra\) is prime. If \(Ra\) is a prime ideal and \(a \mid xy\), then \(xy \in Ra\), so \(x \in Ra\) or \(y \in Ra \iff a \mid x\) or \(a \mid y\).

e) Suppose \(a\) is irreducible. If \(Rx\) is a proper principal ideal of \(R\) with \(Ra \subseteq Rx\), then \(x \mid a\). Since \(a\) has no proper divisors, \(x\) is associated with \(a\) or it is a unit. But \(x\) cannot be a unit because \(Rx \neq R\). So, \(x \sim a\), i.e. \(Rx = Ra\). Suppose now \(Rx\) is maximal among principal proper ideals, and \(d \in R\) is a divisor of \(a\). Then \(Ra \subseteq Rd\), so \(Rd = Ra\) or \(Rd = R\). This means \(d \sim a\) or \(d \sim 1\).

f) If \(a\) is a common divisor of \(x_1, \ldots, x_n\), then \(a \mid r_1x_1 + \ldots + r_nx_n\), for any \(r_1, \ldots, r_n \in R\), so any element in the ideal \(Rx_1 + \ldots + Rx_n\) is divisible by \(a\). The other implication is left to the reader.

g) From f), \(a\) is a common divisor of \(x_1, \ldots, x_n\). Let \(d \in R\) be another common divisor. Because \(a \in Rx_1 + \ldots + Rx_n\), \(\exists c_1, \ldots, c_n \in R\) with \(a = c_1x_1 + \ldots + c_nx_n\). Since \(d \mid x_1, \ldots, d \mid x_n\), we obtain \(d \mid a\).

h), i) are left to the reader.

An essential role in the arithmetic of \(\mathbb{Z}\) is played by the **theorem of division with remainder**: For any \(a, b \in \mathbb{Z}, b \neq 0\), there exist \(q, r \in \mathbb{Z}\), such that \(a = bq + r\) and \(|r| < |b|\) or \(r = 0\).

From this theorem, one deduces other two fundamental properties of \(\mathbb{Z}\):

Every ideal of \(\mathbb{Z}\) is principal (of the form \(n\mathbb{Z}\), for some \(n \in \mathbb{Z}\)).
Any nonzero non-invertible integer can be written as a finite product of prime integers and this writing is unique up to the order of factors and an association of the factors in divisibility. This result is called „the fundamental theorem of integer arithmetic” or „the unique integer factorization theorem”.

Abstracting these properties of \( \mathbb{Z} \), one obtains the notions of Euclidian domain (a domain in which a property analogous to the theorem of division with remainder holds), principal ideal domain (a domain whose every ideal is principal) and, respectively, unique factorization domain (a domain in which every nonzero non-invertible element can be written as a product of primes).

The following sections are devoted to the study of these classes of rings.

I.2 Euclidian domains

2.1 Definition. A domain \( R \) is called an Euclidian domain if there exists a mapping \( \varphi : R \setminus \{0\} \rightarrow \mathbb{N} \) satisfying:

\[
\text{For any } a, b \in R, b \neq 0, \text{ there exist } q, r \in R \text{ such that: } a = bq + r \text{ and } (r = 0 \text{ or } \varphi(r) < \varphi(b)).
\]

We say in this case that \( R \) is an Euclidian domain with respect to \( \varphi \).

The property (DRT) is called the “Division with remainder theorem in \( R \)”; \( q \) is called traditionally the quotient and \( r \) is called the remainder of the division of \( a \) by \( b \).

Of course, the definition above originates in the division with remainder theorems in \( \mathbb{Z} \) (where \( \varphi \) is the absolute value \( \lvert \cdot \rvert : \mathbb{Z} \setminus \{0\} \rightarrow \mathbb{N} \)), respectively in \( K[X] \), with \( K \) a field (where \( \varphi \) is the
degree function $\deg : K[X] \setminus \{0\} \to \mathbb{N})$. These rings are also the most important examples of Euclidian domains.

**2.2 Remark.** Sometimes, in the definition of an Euclidian domain, the following condition on $\varphi$ is also required:

*For any $a, b \in R \setminus \{0\}$, $a \mid b$ implies $\varphi(a) \leq \varphi(b)$.*

This extra condition is not essential and in fact defines the same class of rings as the definition above (see 4.9).

**2.3 Remark.** If $b \in R$, $b \neq 0$, and $\varphi(b) = 0$, then $b \in U(R)$. Indeed, the remainder $r$ of the division of 1 to $b$ satisfies $1 = bq + r$, for some $q \in R$, and $(r = 0$ or $\varphi(r) < \varphi(b) = 0)$. Clearly, the natural number $\varphi(r)$ cannot be $< 0$, so $r = 0$.

The converse is false in general (give a counterexample!).

The Euclidian domains are GCD rings. This follows from the fact that the Euclidian algorithm can be performed in these rings and guarantees the existence of GCD's:

**2.4 Theorem (Euclidian Algorithm).** Let $R$ be an Euclidian domain and let $a, b \in R$, with $b \neq 0$. Then a GCD of $a$ and $b$ exists and it can be found by the following algorithm:

**Algorithm Euclid** ($R, a, b, d$)

**Input:** $a, b \in R$.

**Output:** $d = \text{GCD}(a, b) \in R$.

begin
  if $b = 0$ then $d := a$; **Stop**.
  else
    (**Step 1**) Find $q, r \in R$ with $a = bq + r$ and $(r = 0$ or $\varphi(r) < \varphi(b))$.
    if $r = 0$ then $d := b$; **Stop**.
    else $a := b$, $b := r$; go to **Step 1**
  end
end
Moreover, there exist (and can be algorithmically determined) \( u, v \in R \) such that
\[
d = au + bv.
\]

**Proof.** The algorithm\(^3\) above implies the following sequence of divisions with remainder performed in \( R \):

1. \( a = b q_1 + r_1 \) with \( r_1 = 0 \) or \( \varphi(r_1) < \varphi(b) \);
2. \( b = r_1 q_2 + r_2 \) with \( r_2 = 0 \) or \( \varphi(r_2) < \varphi(r_1) \);
3. \( r_1 = r_2 q_3 + r_3 \) with \( r_3 = 0 \) or \( \varphi(r_3) < \varphi(r_2) \);

\[ \cdots \]
4. \( r_{n-4} = r_{n-3} q_{n-2} + r_{n-2} \) with \( r_{n-2} = 0 \) or \( \varphi(r_{n-2}) < \varphi(r_{n-2}) \);
5. \( r_{n-3} = r_{n-2} q_{n-1} + r_{n-1} \) with \( r_{n-1} = 0 \) or \( \varphi(r_{n-1}) < \varphi(r_{n-2}) \);
6. \( r_{n-2} = r_{n-1} q_n + r_n \) with \( r_n = 0 \).

The existence of the elements \( q_i, r_i \in R \) with the properties above is guaranteed by the condition (1) in the definition of the Euclidean domain. The strictly decreasing sequence of natural numbers \( \varphi(b) > \varphi(r_1) > \varphi(r_2) > \cdots \) must terminate, so there exists \( n \in \mathbb{N}^* \) with \( r_n = 0 \) (which means that the algorithm terminates in a finite number of steps). We must prove that \( r_{n-1} \) (the last nonzero remainder) is a GCD of \( a \) and \( b \).

The relation (n) shows \( r_{n-1} \mid r_{n-2} \). The relation \( (n - 1) \) implies \( r_{n-1} \mid r_{n-3} \). Using relations \( (n - 2) \), \( \cdots \), (3), (2), (1), we obtain (by induction) that \( r_{n-1} \mid b \) and \( r_{n-1} \mid a \). Now let \( e \in R \) be a common divisor of \( a \) and \( b \); then \( e \) also divides \( r_1 = a - bq_1 \). Using (2), \( e \) divides

---

\(^3\) We hope that the algorithm is clear to the reader. We do not want to present a rigorous “pseudo-programming language” or use strict syntax rules from a particular language. Also, this algorithm is intended to serve a theoretical purpose; for instance, “finding” the elements \( q, r \) at step 1 does not imply a description of a concrete procedure (such procedures can be given for particular rings as \( \mathbb{Z}, \mathbb{Q}[X], \ldots \)) and merely uses the fact that these elements exist. Moreover, implementing this algorithm must take into account computer representations of the elements of \( R \), decision algorithms of the equality of two elements in \( R \), addition and multiplication algorithms in \( R \) etc. These important issues are not discussed here.
2.5 Examples. a) \( \mathbb{Z} \) is an Euclidean domain with respect to the absolute value. Here is a proof. Let \( b \in \mathbb{Z}, \, b > 0 \). We shall prove (by induction) that for every \( a \in \mathbb{N} \), there exist \( q, \, r \in \mathbb{N} \) satisfying \( a = bq + r \) and \( r = 0 \) or \( r < b \). If \( a < b, \, q = 0, \, r = a \) satisfy these conditions. If \( a \geq b \), suppose that the claim is true for any \( n < a \) and we prove it for \( a \). Since \( a - b < a \), by induction exist \( q, \, r \in \mathbb{N} \) such that \( a - b = bq + r \) and \( r = 0 \) or \( r < b \). This implies \( a = b(q + 1) + r \) and the claim is proven. If \( a < 0, \, -a \in \mathbb{N} \), so there exist \( q, \, r \in \mathbb{N} \) such that \( -a = bq + r \) and \( r = 0 \) or \( r < b \); so \( a = b(-q) + (-r) \), with \( r = 0 \) or \( |r| < b \). We leave the case \( b < 0 \) to the reader.
Given \( a, b \in \mathbb{Z} \), the quotient \( q \) and remainder \( r \) of the division of \( a \) by \( b \) are not unique: for instance, \( 3 = 2 \cdot 1 + 1 = 2 \cdot 2 + (-1) \). If we require that the remainder should be positive, then \( q \) and \( r \) are unique.

b) The ring \( K[X] \) of polynomials in one indeterminate \( X \) with coefficients in the field \( K \) is Euclidian with respect to \( \text{deg} : K[X] \setminus \{0\} \to \mathbb{N} \). The proof originates in the procedure of long polynomial division taught in school. Let \( f = a_0 + \ldots + a_n X^n \) and \( g = b_0 + \ldots + b_m X^m \in K[X] \), with \( g \neq 0 \) (i.e. \( b_m \neq 0 \)). We prove by induction on \( n = \text{deg} f \). If \( n < m \), set \( q = 0, r = f \). If \( n \geq m \), \( h = f - b_m^{-1} a_n g \) has degree less than \( n \) (the degree \( n \) terms cancel out) and, by induction, there exist \( q, r \in K[X], h = gq + r \), with \( \text{deg} r < m \). Thus, \( f = g(q + b_m^{-1} a_n X^{n-m}) + r \). Note that \( q \) and \( r \) are unique (prove!).

For further reference, we recall:

2.6 Definition. Let \( n \in \mathbb{N} \) be fixed. Two integers \( a, b \in \mathbb{Z} \) are congruent modulo \( n \) if \( n \mid a - b \). We denote this fact by \( a \equiv b \pmod{n} \).

We have \( a \equiv b \pmod{n} \iff a - b \in n\mathbb{Z} \iff a \) and \( b \) yield the same remainder when divided by \( n \). Recall that the relation defined here is an equivalence relation on \( \mathbb{Z} \) and the set of equivalence classes is structured as a ring (which is in fact the factor ring \( \mathbb{Z}/n\mathbb{Z} \)), called the ring of integers modulo \( n \), often denoted by \( \mathbb{Z}_n \).

I.3 Euclidian rings of quadratic integers

Besides \( \mathbb{Z} \) and the polynomial ring in one indeterminate with coefficients in a field, important examples of Euclidian domains are among the rings of the form

\[
\mathbb{Z}[\sqrt{d}] = \{a + b \sqrt{d} \mid a, b \in \mathbb{Z}\}.
\]
A somewhat surprising fact is that the ring \( \mathbb{Z}[\frac{1 + \sqrt{d}}{2}] = \{a + b \frac{1 + \sqrt{d}}{2} | a, b \in \mathbb{Z}\} \) has sometimes better arithmetic properties than \( \mathbb{Z}[\sqrt{d}] \). This fact is closely connected to the theory of the rings of quadratic integers, an important topic in (algebraic) number theory. In the following, some elementary facts on these rings are presented. The unproven statements below are proposed as exercises (some of them in the next chapters). For a systematic treatment of the theory of algebraic integers, see for instance Lang [1964].

3.1. Definition. A subfield of \( \mathbb{C} \) that has dimension 2 viewed as a vector space over \( \mathbb{Q} \) is called a quadratic number field.

Using elementary tools of field extensions theory, one can readily show that any quadratic number field is of the form \( \mathbb{Q}[\sqrt{d}] = \{a + b \sqrt{d} | a, b \in \mathbb{Q}\} \), where \( d \in \mathbb{Z} \) is squarefree.

3.2. Definition. A complex number that is the root of a monic polynomial in \( \mathbb{Z}[X] \) is called integral over \( \mathbb{Z} \) (or algebraic integer). For instance, \( \sqrt{2} \) is integral over \( \mathbb{Z} \), but \( \frac{1}{2} \) is not. Sometimes, for avoiding confusions, the numbers in \( \mathbb{Z} \) are called rational integers.\(^5\)

An element of a quadratic number field that is integral over \( \mathbb{Z} \) is called a quadratic integer. One can prove that: a quadratic integer is a root of a monic polynomial in \( \mathbb{Z}[X] \) of degree 2.

Fix \( d \in \mathbb{Z} \), squarefree. If \( \alpha = a + b \sqrt{d} \in \mathbb{Q}[\sqrt{d}] \), the element \( \bar{\alpha} = a - b \sqrt{d} \) is called the conjugate of \( \alpha \). Define the norm \( N : \mathbb{Q}[\sqrt{d}] \to \mathbb{Q} \) and the trace \( Tr : \mathbb{Q}[\sqrt{d}] \to \mathbb{Q} \),

\[
N(\alpha) := \alpha \bar{\alpha} = a^2 - db^2
\]

\(^4\) A polynomial is called monic if the coefficient of its monomial of maximum degree is 1.

\(^5\) Because every algebraic integer over \( \mathbb{Z} \) which is rational (in \( \mathbb{Q} \)) must be in \( \mathbb{Z} \). Prove!
I. Arithmetic in integral domains

\[ \text{Tr}(\alpha) := \alpha + \bar{\alpha} = 2a \]

for any \( \alpha = a + b \sqrt{d} \in \mathbb{Q}[\sqrt{d}] \) \((a, b \in \mathbb{Q})\).

The norm \( N \) is multiplicative and the trace \( \text{Tr} \) is additive: for any \( \alpha, \beta \in \mathbb{Q}[\sqrt{d}] \),

\[ N(\alpha \beta) = N(\alpha)N(\beta), \quad \text{Tr}(\alpha + \beta) = \text{Tr}(\alpha) + \text{Tr}(\beta). \]

One can prove that, for any \( x \in \mathbb{Q}[\sqrt{d}] \): \( x \) is integral over \( \mathbb{Z} \) \((x \text{ is a quadratic integer}) \) \( \iff \) \( \text{Tr}(x) \in \mathbb{Z} \) and \( N(x) \in \mathbb{Z} \).

The quadratic integers in \( \mathbb{Q}[\sqrt{d}] \) form a ring, called the ring of integers of \( \mathbb{Q}[\sqrt{d}] \). This ring is traditionally called a ring of quadratic integers (imaginary if \( d < 0 \), respectively real if \( d > 0 \)). We have the following (for the proof, see the Exercises):

3.3 Proposition. The ring of integers of \( \mathbb{Q}[\sqrt{d}] \) is \( \mathbb{Z}[\theta] \), where \( \mathbb{Z}[\theta] = \{ a + b \theta \mid a, b \in \mathbb{Z} \} \) and

\[ \theta = \begin{cases} \sqrt{d} & \text{if } d \equiv 2 \text{ or } 3 \pmod{4} \smallskip \medskip (0) \smallskip \medskip (0) \smallskip \medskip 21 \smallskip \medskip 21 \smallskip \medskip d \equiv 1 \pmod{4} \end{cases}. \]

In what follows, for a fixed \( d \in \mathbb{Z} \), squarefree, \( \theta \) denotes the number above.

We remark also that

\[ \mathbb{Q}[\sqrt{d}] = \mathbb{Q}[\theta] = \{ a + b \theta \mid a, b \in \mathbb{Q} \}. \]

If \( d \equiv 1 \pmod{4} \), then \( \mathbb{Z}[\theta] = \{ a + b (1 + \sqrt{d})/2 \mid a, b \in \mathbb{Z} \} \) can also be described as the set of complex numbers of the form \((u + v \sqrt{d})/2\), with \( u, v \in \mathbb{Z} \) having the same parity.

According to the above, \( \mathbb{Z}[i], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[1 + i \sqrt{3})/2], \mathbb{Z}[1 + \sqrt{5})/2] \) are examples of rings of quadratic integers.

The norm \( N : \mathbb{Q}[\sqrt{d}] \to \mathbb{Q} \) has the property that \( N(\alpha) \in \mathbb{Z}, \forall \alpha \in \mathbb{Z}[\theta] \), as we saw above. We obtain a mapping:

\[ |N| : \mathbb{Z}[\theta] \to \mathbb{N}, \; |N|(\alpha) = |N(\alpha)|, \forall \alpha \in \mathbb{Z}[\theta]. \]

A natural problem arises:
For which \( d \in \mathbb{Z}, \) squarefree, \( \mathbb{Z}[\theta] \) is Euclidian with respect to \( |N|? \)

We remark first that, since the norm \( N : \mathbb{Q}[\sqrt{d}] \rightarrow \mathbb{Q} \) is multiplicative \( (N(xy) = N(x)N(y), \forall x, y \in \mathbb{Q}[\sqrt{d}]) \), \( |N| : \mathbb{Z}[\theta] \rightarrow \mathbb{N} \) is also multiplicative.

If \( d < 0 \), the representation of the complex numbers in \( \mathbb{Z}[\theta] \) in the plane yields a grid (rectangular if \( d \equiv 2 \) or \( 3 \) (mod 4) and oblique if \( d \equiv 1 \) (mod 4)); moreover, for any \( x, y \in \mathbb{Q}[\sqrt{d}] \), \( |N|(x - y) \) is the Euclidian distance between the points \( x \) and \( y \).

If \( d > 0 \), \( \mathbb{Z}[\theta] \) is a (dense) subset of \( \mathbb{R} \) and no geometric interpretation is available.

**3.4 Lemma.** Let \( d \in \mathbb{Z}, \) squarefree. Then \( R := \mathbb{Z}[\theta] \) is Euclidian with respect to \( |N| \) if and only if for any \( x \in \mathbb{Q}[\sqrt{d}] \), there exists \( \gamma \in R \) such that \( |N|(x - \gamma) < 1 \).

**Proof.** \( \Rightarrow \) \( \mathbb{Q}[\sqrt{d}] \) is the field of quotients of \( R \), i.e.: \( \forall x \in \mathbb{Q}[\sqrt{d}], \exists \alpha, \beta \in R, \beta \neq 0 \) such that \( x = \alpha/\beta \). Since \( R \) is Euclidian with respect to \( |N| \), \( \exists \gamma, \delta \in R \) such that \( \alpha = \beta \gamma + \delta \), with \( |N|(\delta) < |N|(\beta) \) or \( \delta = 0 \). So,

\[
x = \alpha/\beta = \gamma + \delta/\beta,
\]

with \( |N|(x - \gamma) = |N|(\delta/\beta) = |N|(\delta)/(\beta) < 1 \).
“⇐” Let $\alpha, \beta \in R$. For $x = \alpha/\beta \in \mathbb{Q}[^\sqrt{d}]$, $\exists \gamma \in R$ with $|N|(x - \gamma) < 1$. Let $\delta = x - \gamma \in \mathbb{Q}[^\sqrt{d}]$. Thus, $\alpha = \beta\gamma + \beta\delta$, with $\beta\delta = \alpha - \beta\gamma \in R$ and $|N|(\beta\delta) = |N|(|\beta|N|\delta|) < 1$. □

If $d < 0$, because $\mathbb{Q}[^\sqrt{d}]$ is dense\(^6\) in $\mathbb{C}$, the lemma can be rephrased as follows:

The ring $\mathbb{Z}[\theta]$ is Euclidian with respect to $|N|$ if and only if any point in the complex plane is situated at a distance smaller than 1 to some point in the grid $\mathbb{Z}[\theta]$.

3.5 Proposition\(^7\). Let $d \in \mathbb{Z}, d < 0$, squarefree.

a) If $d$ is congruent with 2 or 3 (mod 4), $\mathbb{Z}[^\sqrt{d}]$ is Euclidian with respect to $|N|$ if and only if $d = -1$ or $d = -2$.

b) If $d \equiv 1$ (mod 4), $\mathbb{Z}[t(1 + \sqrt{d})/2]$ is Euclidian with respect to $|N|$ if and only if $d \in \{-3, -7, -11\}$.

Proof. a) $\mathbb{Z}[^\sqrt{d}]$ is Euclidian with respect to $|N|$ if and only if all points inside a rectangle of the grid are situated at a distance less than 1 from some vertex of the rectangle. The greatest distance to the vertices is attained at the intersection of the diagonals, at the distance of $\sqrt{1 - d}/2$ to any vertex. We have $\sqrt{1 - d}/2 < 1$ if and only if $d > -3$, i.e. $d = -1$ or $d = -2$.

b) In this case, the grid forms isosceles triangles with base 1 and height $\sqrt{-d}/2$. The points inside the triangle have distance less than 1 to some vertex if and only if the circles of radius 1 centered in base vertices intersect in a point $P$ situated at a distance less than 1 to the third vertex. An easy geometric reasoning shows that the distance from $P$ to the third vertex is $(\sqrt{-d} - \sqrt{3})/2$. We must have then $(\sqrt{-d} - \sqrt{3})/2 < 1 \iff d > -7 - 4\sqrt{3} \iff d \in \{-3, -7, -11\}$. □

\(^6\)For any $z \in \mathbb{C}$ and any $\varepsilon \in \mathbb{R}, \varepsilon > 0$, there exists $x \in \mathbb{Q}[^\sqrt{d}]$ such that $|z - x| < \varepsilon$.

\(^7\)This result was obtained in 1923 by L. E. Dickson.
We obtain that the following rings are Euclidian with respect to $|N|$: $\mathbb{Z}[i], \mathbb{Z}[i\sqrt{2}], \mathbb{Z}[\frac{1+i\sqrt{3}}{2}], \mathbb{Z}[\frac{1+i\sqrt{7}}{2}], \mathbb{Z}[\frac{1+i\sqrt{11}}{2}]$.

One can prove that these are all the imaginary Euclidian rings of quadratic integers (not necessarily with respect to $|N|$). The real case $d > 0$ has no geometric interpretation and is considerably more difficult.

I.4 Principal ideal domains

4.1 Definition. A domain $R$ is called a principal ideal domain (PID) if any ideal of $R$ is principal. In other words, for every ideal $I$ of $R$, there exists $a \in R$ such that $I = Ra$.

Any field is a PID\textsuperscript{8}. The most important examples of PIDs are given by the following result.

4.2 Theorem. Any Euclidian domain is a principal ideal domain.

Proof. Let $R$ be Euclidian with respect to $\varphi$ and let $I$ be a nonzero ideal of $R$. The set of natural numbers $\{\varphi(x) \mid x \in I, x \neq 0\}$ contains a minimum element: $\exists a \in I, a \neq 0$, such that $\varphi(a) = \min\{\varphi(x) \mid x \in I, x \neq 0\}$ ($a$ may not be unique). We claim that $a$ is a generator of the ideal $I$. Obviously, $Ra \subseteq I$. In order to prove the opposite inclusion, suppose by contradiction that there exists $b \in I \setminus Ra$. Applying the division with remainder in $R$, we obtain elements $q, r \in R$ such that $b = aq + r, r \neq 0$ (since $b \notin Ra$) and $\varphi(r) < \varphi(a)$. But $a, b \in I$, so $r \in I$ and $\varphi(r) < \varphi(a)$ contradict the choice of $a$.

\textsuperscript{8}Sometimes the fields are not considered principal ideal domains by definition.
For instance, if $K$ is a field, $K[X]$ is a PID; if $I \neq 0$ is an ideal in $K[X]$, a generator of $I$ is a polynomial $g \in I$ of minimum degree among the degrees of nonzero polynomials in $I$.

The principal ideal domains are GCD-domains; for any $a, b \in R$, there exists their GCD, namely any generator of the ideal $aR + bR$:

**4.3 Proposition.** Let $R$ be a PID and let $a, b \in R$. Then:

a) A GCD $d$ of $a$ and $b$ exists and, for some $u, v \in R$, $d = au + bv$. Moreover, the element $d \in R$ is a GCD of $a$ and $b$ if and only if $dR = aR + bR$.

b) The element $m \in R$ is a LCM of $a$ and $b$ if and only if $mR = aR \cap bR$.

**Proof.**

a) Since $R$ is a PID, there exist $d \in R$ such that the ideal $aR + bR = \{ax + by \mid x, y \in R\}$ is generated by $d$. Since $d \in aR + bR$, there exist $u, v \in R$ such that $d = au + bv$. We have $a, b \in dR$, so $d \mid a, d \mid b$. If $e \in R$ is such that $e \mid a, e \mid b$, then $e \mid ax + by$, $\forall x, y \in R$. In particular, $e \mid d$. Thus, any generator $d$ of $aR + bR$ is a GCD of $a$ and $b$.

Conversely, if $d$ is a GCD of $a$ and $b$, $d$ divides $a$ and $b$, so $dR \supseteq aR$ and $dR \supseteq bR$. Thus, $dR \supseteq aR + bR$. If $e$ is a generator of $aR + bR$, this means $d \mid e$. But $e$ is also a common divisor of $a$ and $b$, so $d$ is associated to $e$. Thus, $dR = eR = aR + bR$.

b) The proof is similar to the one above and is proposed as an exercise.

Recall that, for $a, b \in R$, the notation $(a, b)$ is used to designate the ideal generated by $a$ and $b$, $aR + bR$. The proposition above shows that this notation is consistent to the fact that it designates also the GCD of $a$ and $b$ (which is a generator of the ideal $aR + bR$).

**4.4 Example.** Let $R$ be a domain that is not a field and take $r \in R$ nonzero and non invertible. Then the ideal $(r, X)$ in $R[X]$ is not princi-
pal, so \( R[X] \) is not a PID. In particular, the rings \( \mathbb{Z}[X] \), \( K[X, Y] \), with \( K \) a field, are not PID's.

Indeed, suppose that for some \( f \in R[X] \), we have \( (f) = (r, X) \). Then \( f \mid r \). We obtain \( \deg f = 0 \), so \( f \in R \). Since \( f \mid X \), there exists \( g \in R[X] \) with \( X = fg \), so \( f \) is invertible in \( R \). Thus, the GCD of \( r \) and \( X \) is 1, so \( f = 1 \). But the ideal generated by \( r \) and \( X \) does not contain 1: if \( h, q \in R[X] \) are such that \( 1 = hr + qX \), setting \( X = 0 \) in this equality of polynomials, it follows that \( 1 = h(0) \cdot r \), which means \( r \) is invertible, a contradiction.

Using Proposition 1.19 and the fact that any PID is a GCD-domain, we obtain:

**4.5 Proposition.** In a PID, an element is irreducible if and only if it is a prime element. \( \square \)

Thus, the domains containing irreducibles that are not primes are not principal ideal domains. such a ring is \( \mathbb{Z}[\sqrt{-5}] \), as Example 1.18 shows.

**4.6 Corollary.** In a PID \( R \), the prime nonzero ideals are maximal ideals. Any maximal ideal is of the form \( pR \), where \( p \) is irreducible in \( R \). Moreover, \( p \in R \) is irreducible if and only if \( pR \) is a maximal ideal.

**Proof.** It is sufficient to remark that any nonzero prime ideal is principal, generated by a prime \( p \) (Prop. 1.20,d)). The element \( p \) is irreducible, so (Prop. 1.20,e)) the ideal \( pR \) is maximal. The other statements are obvious. \( \square \)

The case \( R = \mathbb{Z} \) of the following proposition is known as the “the fundamental theorem of integer arithmetic”. Recall that

\[
R^\circ = \{ x \in R \mid x \text{ is nonzero, non-invertible} \}.
\]
4.7 Theorem. Let $R$ be a PID. Then every nonzero non-invertible element $r$ in $R$ is a finite product of prime elements.\footnote{The products may contain a single factor (i.e., the element itself is a prime).}

Proof. Since $R$ is a PID, the primes in $R$ coincide with the irreducibles in $R$. Suppose by contradiction that there exists $r_0 \in R^\circ$ such that $r_0$ cannot be written as a finite product of irreducibles. In particular, $r_0$ is reducible, so $r_0 = r_1s_1$, with $r_1, s_1 \in R^\circ$, not associated with $r_0$. If $r_1$ and $s_1$ are finite products of irreducibles, then $r_0$ is too, false. So, at least one of them (say $r_1$) is not a product of irreducibles. We obtain thus $r_1 \in R^\circ$ with $r_1 \mid r$, $r_1 \sim r$ and $r_1$ is not a product of irreducibles. This reasoning applies to $r_1$, so we get $r_2 \in R^\circ$, $r_2 \mid r_1$, $r_2 \sim r_1$, and $r_2$ is not a product of irreducibles. By induction, there exists a sequence $(r_n)_{n \geq 0}$ of elements in $R$, with the property that for any $n \in \mathbb{N}$, $r_{n+1}$ is a proper divisor of $r_n$. In other words, we obtain a strictly increasing sequence of ideals $r_0 R \subset r_1 R \subset \ldots \subset r_n R \subset r_{n+1} R \subset \ldots$. This is impossible in a PID, as the following lemma shows.

4.8 Lemma. Let $R$ be a PID and let $(I_n)_{n \geq 0}$ be a sequence of ideals in $R$ such that $I_n \subseteq I_{n+1}$, for any $n \in \mathbb{N}$. Then there exists $m \in \mathbb{N}$ such that $I_m = I_{m+i}$, for any $i \in \mathbb{N}$. (In other words, every ascending sequence of ideals is stationary).

Proof. Let $I$ be the union of all ideals $I_n$, $n \in \mathbb{N}$. Since the sequence $(I_n)_{n \geq 0}$ is ascending, $I$ is an ideal in $R$: if $x, y \in I$, then, for some $i, j \in \mathbb{N}$, $x \in I_i$, $y \in I_j$. So $x, y \in I_t$, where $t = \max(i, j)$, hence $x + y \in I_t \subseteq I$. If $r \in R$, $rx \in I_t \subseteq I$. But $R$ is a PID, so there exists $a \in R$ such that $I = ar$. Since $a \in I$, there exists $m \in \mathbb{N}$ such that $a \in I_m$, so $ar \subseteq I_m \subseteq I = ar$. So $ar = I = I_{m+i}$, $\forall i \in \mathbb{N}$.  \hfill \QED

A ring $R$ satisfying the ascending chain condition (ACC): every ascending sequence of ideals of $R$, $I_0 \subseteq I_1 \subseteq \ldots$, is stationary (there ex-
ists $m \in \mathbb{N}$ such that $I_m = I_{m+i}, \forall i \in \mathbb{N}$) is called a Noetherian ring.\footnote{Emmy Noether (1882–1935), German.} We have shown that every PID is Noetherian. Note that the proof above can be slightly modified to obtain the theorem: A ring in which every ideal is finitely generated is Noetherian. The converse is also true. Noetherian rings are an important topic in Commutative Algebra.

We have remarked that sometimes in the definition of the Euclidian domain $R$ with respect to $\varphi : R \setminus \{0\} \to \mathbb{N}$, an extra condition is required: for any $a, b \in R^*$, $a \mid b$ implies $\varphi(a) \leq \varphi(b)$. We show that this condition is not essential.

4.9 Proposition. Let $R$ be a Euclidian domain with respect to $\varphi : R^* \to \mathbb{N}$. Then there exists $\psi : R^* \to \mathbb{N}$ such that $R$ is Euclidian with respect to $\psi$ and, for any $a, b \in R^*$, $a \mid b$ implies $\psi(a) \leq \psi(b)$.

Proof. Let $\psi : R^* \to \mathbb{N}$, defined by $\psi(x) = \min\{\varphi(y) \mid y \sim x\}, \forall x \in R$. Clearly, $x \sim y$ implies $\psi(x) = \psi(y)$. We claim that $R$ is Euclidian with respect to $\psi$. Let $a, b \in R$, $b \neq 0$ and let $b_0$ be associated with $b$ such that $\psi(b) = \varphi(b_0)$. There exist $q, r \in R$ such that $a = b_0q + r$ and $r = 0$ or $\varphi(r) < \varphi(b_0)$. Since $b_0 = bu$, for some unit $u$, we have $a = b(uq) + r$ and $r = 0$ or $\psi(r) \leq \varphi(r) < \varphi(b_0) = \psi(b)$. Now let $a, b \in R^*$ with $a \mid b$. As seen in the proof of 4.2, a generator of the ideal $aR$ is an element $c$ (which must be associated with $a$) such that $\psi(c) = \min\{\psi(x) \mid x \in aR, x \neq 0\}$. So $\psi(a) = \psi(c) \leq \psi(b)$, since $b \in aR$. \square

Are there any principal ideal domains that are not Euclidian? The answer is affirmative, but the examples are not easy to find. Such an example is $\mathbb{Z}\left[\frac{1+i\sqrt{19}}{2}\right]$, as shown by Dedekind in 1894.
I.5 Unique factorization domains

5.1 Definition. A domain $R$ with the property that every nonzero and non-invertible element is a finite product\(^{11}\) of prime elements is called a *unique factorization domain* (UFD).

Theorem 4.7 shows that any principal ideal domain (thus also any Euclidian domain) is a UFD. Every field is a UFD, because it has no nonzero and non-invertible elements.

5.2 Proposition. Every irreducible element in a UFD is also a prime.

Proof. Let $R$ be a UFD and let $p$ be irreducible in $R$. Since $p \in R^\circ$, $p$ is a product of primes. But this product can have only one factor, otherwise $p$ would not be irreducible. So, $p$ is itself a prime. \(\square\)

The next Proposition justifies and explains the epithet “unique” in the name of a UFD.

5.3 Proposition. Let $R$ be a domain and $r \in R^\circ$. If $r$ has a prime decomposition, then this decomposition is unique up to an ordering of the factors and an association in divisibility. This means that: if $r = p_1 \ldots p_n = q_1 \ldots q_m$ are two prime decompositions of $r$, then $m = n$ and there exists a permutation $\sigma$ of the set $\{1, \ldots, n\}$ such that $p_i \sim q_{\sigma(i)} \forall i \in \{1, \ldots, n\}$.

Proof. Let $r = p_1 \ldots p_n$ be a prime decomposition of $r$. Call the natural number $n$ the *length* of the decomposition. Define $l(r)$ as the smallest $n \in \mathbb{N}^*$ such that there exists a prime decomposition of $r$ of length $n$.

\(^{11}\) Such a product is also called a *prime decomposition* of the element. Products may have a single factor.
We prove the claim by induction on \( l(r) \).

If \( l(r) = 1 \), then let \( r = p_1 = q_1 \ldots q_m \), with \( p_1, q_1, \ldots, q_m \) prime. Since \( r \) is a prime and \( r \) divides the product \( q_1 \ldots q_m \), \( r \) divides one of the factors, say \( q_1 \) (relabel if necessary). But \( q_1 \) is irreducible, so we have \( r \sim q_1 \). Thus \( r = q_1 u \), with \( u \) invertible. We must prove that \( m = 1 \). If \( m \geq 2 \), simplifying by \( q_1 \) in \( q_1 u = q_1 \ldots q_m \), we obtain \( q_2 \ldots q_m = u \), so \( q_2, \ldots, q_m \) are invertible, contradiction.

Suppose the claim is true for any \( x \in R^o \) with \( l(x) < n \). Let \( r \in R \), \( r = p_1 \ldots p_n = q_1 \ldots q_m \), with \( p_1, \ldots, p_n, q_1, \ldots, q_m \) primes. Because \( p_n \) is prime, \( \exists i \in \{1, \ldots, n\} \) such that \( p_n \mid q_i \). But \( q_i \) is irreducible, so \( p_n \sim q_i \), that is, \( v p_n = q_i \), with \( v \) a unit. Simplifying by \( p_n \), we obtain the relation \( p_1 \ldots p_{n-1} = v q_1 \ldots q_{i-1} q_{i+1} \ldots q_m \). The induction hypothesis applied to \( p_1 \ldots p_{n-1} \) shows that \( n - 1 = m - 1 \) and \( p_1, \ldots, p_{n-1} \) are associated with \( q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_m \), possibly in other order. \( \square \)

A UFD is also a GCD-domain. This is an important property of the UFD's. Several other characterizations are collected in the next theorem.

**5.4 Theorem.** Let \( R \) be a domain. The following statements are equivalent:

a) \( R \) is a UFD.

b) Every element in \( R^o \) is a finite product of irreducibles and every irreducible is a prime.

c) Every element in \( R^o \) has a decomposition in irreducible factors, unique up to a reordering of the factors and an association in divisibility.

d) Every element in \( R^o \) has a decomposition in irreducible factors and every two elements have a GCD.

**Proof.**

a)\( \Rightarrow \)b) It follows from Prop. 5.2.

b)\( \Rightarrow \)c) It follows from Prop. 5.3.
c) ⇒ d) Let \( a, b \in R^\circ \) (if \( a, b \in \{0\} \cup U(R) \), it is trivial to exhibit a GCD of \( a \) and \( b \)). A GCD of \( a \) and \( b \) can be determined by “taking the common prime factors, at the smallest power”. To be precise, let \( P \) be a system of representatives of the equivalence classes of the irreducible elements in \( R \) (with respect to the association in divisibility). This means that every irreducible in \( R \) is associated to exactly one element in \( P \). Property \( c \) assures that there exist and are unique distinct \( p_1, \ldots, p_n \in P, s_1, \ldots, s_n, t_1, \ldots, t_n \in \mathbb{N}, u, v \in U(R) \) such that
\[
a = p_1^{s_1} \cdots p_n^{s_n} u \quad \text{and} \quad b = p_1^{t_1} \cdots p_n^{t_n} v.
\]

The uniqueness claim follows the uniqueness assertion in \( c \). Let \( r_i = \min(s_i, t_i) \) and define \( d = p_1^{r_1} \cdots p_n^{r_n} \). We have \( d \mid a, d \mid b \). If \( e \mid a, e \mid b \), then every irreducible \( c \in P \) that divides \( e \) divides also \( a \) and \( b \). This implies \( c \in \{p_1, \ldots, p_n\} \). Indeed, otherwise \( a \) (or \( b \)) would possess two decompositions in irreducible factors, one containing \( c \) and the other one not, contradicting the uniqueness property. So \( e = p_1^{w_1} \cdots p_n^{w_n} q \), with \( w_1, \ldots, w_n \in \mathbb{N}, q \in U(R) \). From \( e \mid a \) it follows that \( w_i \leq s_i \), and \( e \mid b \) implies \( w_i \leq t_i, i = 1, \ldots, n \). So \( w_i \leq r_i \) and \( e \mid d \).

\( d) \Rightarrow a) \) Prop. 1.19 makes sure that any irreducible is a prime, since \( R \) is a GCD domain. The implication is now obvious.

Note that in a UFD any two elements \( a \) and \( b \) have a LCM. Given two prime decompositions of \( a \) and \( b \), their LCM can be determined by taking “all prime factors in the decompositions, at the greatest exponent”; with the notations in the proof above, define \( q_i = \max(s_i, t_i) \); the element \( m = p_1^{q_1} \cdots p_n^{q_n} \) is then a LCM of \( a \) and \( b \).

5.5 Example. \( \mathbb{Z}[\sqrt{-5}] \) is not a UFD, since \( 2 \) is irreducible and not a prime (cf. Ex. 1.18). The reader can also check that \( 6 \) has two decompositions as a product of irreducibles in \( \mathbb{Z}[\sqrt{-5}] \):
\[
6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}),
\]
and \( 2 \) is not associated with \( 1 + \sqrt{-5} \) or with \( 1 - \sqrt{-5} \).
5.6 Proposition. Let $R$ be a UFD, $n \in \mathbb{N}^*$ and $a, b_1, \ldots, b_n \in R$. If $(a, b_i) = 1$, for any $1 \leq i \leq n$, then $(a, b_1 \ldots b_n) = 1$.

Proof. It is enough to show that there is no prime $p$ that simultaneously divides $a$ and $b_1 \ldots b_n$. If $p$ is such an element, then there exists $j, 1 \leq j \leq n$ such that $p \mid b_j$. Since $p \mid a$, we have $p \mid (a, b_j) = 1$. Therefore, $p$ is invertible, contradiction.

The UFDs have many characterizations. The following, due to Kaplansky, is inspired from Commutative Algebra techniques.

5.7 Theorem. Let $R$ be a domain. Then $R$ is a UFD if and only if any nonzero prime ideal of $R$ contains a prime element.

Proof. Suppose $R$ is a UFD and $P \neq 0$ is a prime ideal. Take $a \in P$, $a \neq 0$. Since $P \neq R$, $a$ is not a unit, so $a$ decomposes in prime factors: $a = p_1 \ldots p_n$. Because $P$ is a prime ideal, there exists $i$ such that $p_i \in P$.

Suppose now that any nonzero prime ideal in $R$ contains a prime element. Consider

$$S = \{ a \in R \mid a \in U(R) \text{ or } a \text{ is a product of prime elements} \}.$$ 

If $S = R - \{0\}$, then $R$ is a UFD. Suppose, by contradiction, that there exists $a \in R$, nonzero, with $a \notin S$. We note that $S$ is a multiplicative system ($1 \in S$ and $\forall a, b \in S$, $ab \in S$). Then there exists a prime ideal $P$ in $R$ containing $a$, with $P \cap S = \emptyset$. This fact, together with the hypothesis, implies the existence of a prime element $p \in P$. But $p \in S$, contradicting $P \cap S = \emptyset$. So we must have $S = R - \{0\}$.

We have to prove the existence of $P$. We will prove that any ideal $I$, maximal with the properties $I \cap S = \emptyset$ and $aR \subseteq I$, is prime. Let us prove first that such an ideal exists. The idea is to use Zorn's Lemma. Consider the set of ideals (ordered by inclusion):

$$J = \{ I \text{ ideal in } R \mid I \cap S = \emptyset \text{ and } aR \subseteq I \}.$$ 

First, $J$ is not empty, since $aR \in J$. Indeed, if $ar \in S$ for some $r \in R$, then $ar$ is a unit or is a product of primes. If $ar$ is a unit, then $a$ is also a unit, contradiction. If $ar$ is a product of primes, we show that
a ∈ S, by induction on the number of factors. If ar = p, with p a prime implies p ∼ r (so a is a unit) or p ∼ a (so a ∈ S). If ar = p₁ ... pₙ with p₁, ..., pₙ primes, then p₁| a or p₁| r. If p₁| a, then a = bp₁; simplifying, br = p₂ ... pₙ. The induction hypothesis shows that b ∈ S, so a ∈ S. If p₁| r, let r = cp₁, for some c, so ac = p₂ ... pₙ. Again by induction, a ∈ S.

J is inductively ordered: every chain in J, indexed by some set L, (Iᵢ)ᵢ∈L, is bounded above in J by ∪ᵢ∈L Iᵢ. The checking is straightforward and is left to the reader.

Zorn's Lemma guarantees now that some maximal element P ∈ J exists.

The ideal P is prime: let x, y ∈ R, with x ∉ P and y ∉ P. If, by absurd, xy ∈ P, consider the ideals P + Rx and P + Ry, which strictly include P. Since P is maximal, there exist some elements s ∈ S ∩ (P + Rx) and t ∈ S ∩ (P + Ry); then st ∈ S ∩ P, contradiction. □

The following important result shows the property of R being a factorization domain is inherited by the polynomial ring R[X]. This has far from trivial consequences: for instance, it is not obvious that every polynomial in n indeterminates with coefficients in Z decomposes uniquely in irreducible factors.

5.8 Theorem. If R is a UFD, then the polynomial ring in one indeterminate R[X] is a UFD.

The proof of this theorem needs some preparations, which are also interesting on their own. First, we determine the units of R[X].

5.9 Proposition. Let R be a domain. Then

U(R[X]) = U(R) and R[X]° = R° ∪ \{f ∈ R[X] | \deg f ≥ 1\}.

In particular, if K is a field, U(K[X]) = K° and

K[X]° = \{f ∈ K[X] | \deg f ≥ 1\}.
Proof. $U(R) \subseteq U(R[X])$ is evident. If $f$ is a unit $R[X]$, there is some $g$ such that $fg = 1$. Since $R$ is a domain, $\deg f + \deg g = 0$, so $\deg f = 0 = \deg g$, which means that $f, g \in R$. From $fg = 1$ we deduce that $f \in U(R)$.

If $K$ is a field, then $U(K) = K^*$ and the other claims follow.

5.10 Definition. Let $R$ be a UFD and let $f = a_0 + a_1X + \ldots + a_nX^n \in R[X]$. The GCD of the coefficients $a_0, a_1, \ldots, a_n$ is called the content of $f$, denoted $c(f)$. A polynomial with content (associated with) 1 is called a primitive polynomial. Note that $f$ is primitive if and only if there exists no prime $p$ in $R$ such that $p$ divides all the coefficients of $f$. Any $f \in R[X]$ can be written

$f = c(f) \cdot g$, for some primitive $g \in R[X]$.

Also, if $f = a \cdot g$, with $a \in R$ and $g$ primitive, then $c(f) = a$.

5.11 Proposition. a) Let $R$ be a domain. If $p$ is prime in $R$, then $p$ is prime in $R[X]$, too.

b) Let $R$ be a UFD and let $f, g$ be primitive in $R[X]$. Then the product $fg$ is also primitive.\footnote{This is called “Gauss' Lemma”.}

c) Let $R$ be a UFD and let $f, g \in R[X]$. Then $c(fg) = c(f) \cdot c(g)$.

Proof. a) Note first that: $p$ divides a polynomial in $R[X]$ if and only if $p$ divides all the coefficients of the polynomial. Let $f = a_0 + a_1X + \ldots + a_nX^n$, $g = b_0 + b_1X + \ldots + b_mX^m \in R[X]$ such that $p \nmid f$ and $p \nmid g$. Let us prove that $p \nmid fg$. From $p \nmid f$ we deduce that there exists $i, 0 \leq i \leq n$, such that $p \nmid a_i$. Take $i$ minimal with $p \nmid a_i$. Similarly, let $j$ be minimal such that $p \nmid b_j$. Then the coefficient of $X^{i+j}$ in $fg$ is

$$\sum_{k+l=i+j} a_kb_l$$

In this sum, $a_kb_l$ is not divisible by $p$ and the other terms are divisible by $p$ (as products of two factors, at least one of which is divisible.
by \( p \). So, the coefficient of \( X^{i+j} \) is not divisible by \( p \) and neither is the polynomial \( fg \).

b) If \( fg \) is not primitive, there exists \( p \in R \), prime, such that \( p \mid fg \). The previous paragraph implies \( p \mid f \) or \( p \mid g \), contradiction.

c) Let \( f = c(f) \cdot f_1, g = c(g) \cdot g_1 \), where \( f_1 \) and \( g_1 \) are primitive. Then
\[
fg = c(f)c(g) \cdot f_1g_1,
\]
with \( f_1g_1 \) primitive by b). It is clear now that \( c(fg) = c(f)c(g) \). \( \square \)

5.12 Proposition. Let \( R \) be a UFD and let \( K \) be its field of quotients. Then a polynomial \( f \) in \( R[X] \) of degree \( \geq 1 \) is irreducible in \( R[X] \) if and only if \( f \) is primitive and irreducible in \( K[X] \).

Proof. Let \( f \) be irreducible in \( R[X] \). Then \( f \) is clearly primitive. Let us prove that \( f \) is irreducible in \( K[X] \). If \( f = gh \), with \( g, h \in K[X] \), by multiplying with the LCM of the denominators of the coefficients of \( g \) and \( h \), we get something like \( af = g_1h_1 \), for some polynomials \( g_1, h_1 \in R[X] \) and some \( a \in R \). Computing the contents, we have \( a = c(g_1)c(h_1) \), since \( c(f) = 1 \). We have \( g_1 = c(g_1) \cdot g_2, h_1 = c(h_1) \cdot h_2 \), with primitive \( g_2, h_2 \in R[X] \). So \( af = c(g_1) \cdot c(h_1) \cdot g_2 \cdot h_2 \); simplifying by \( a = c(g_1)c(h_1) \), we obtain \( f = g_2h_2 \). Since \( f \) is irreducible in \( R[X] \), \( \deg g_2 = 0 \) or \( \deg h_2 = 0 \). But \( \deg g = \deg g_1 = \deg g_2 \) and similarly for \( h \), so \( \deg g = 0 \) or \( \deg h = 0 \).

Conversely, if \( f \) is irreducible in \( K[X] \), it has no proper divisors (of degree \( \geq 1 \)) in \( K[X] \); so it does not have divisors of degree \( \geq 1 \) in \( R[X] \). Since \( f \) is primitive, \( f \) has no non-invertible factors of degree 0. \( \square \)

This result is also important from a practical point of view: in order to prove that some polynomial with coefficients in \( R \) is irreducible in \( K[X] \), it is sufficient to prove it is irreducible in \( R[X] \), which is often easier to accomplish.

We can give now the Proof of the Theorem 5.8. We will use the characterization in 5.4.b). Let \( R \) be a UFD, \( K \) its field of quotients and \( f \in R[X] \), irreducible. Let us prove that \( f \) is a prime. If \( f \mid gh \), with \( g, h \in \text{证明} \)
$R[X]$, because $f$ is irreducible in $K[X]$ (and thus also prime in $K[X]$), we have $f \mid g$ or $f \mid h$ in $K[X]$. Say $f \mid g$ in $K[X]$; so there exist $a \in R, q \in R[X]$, such that $ag = fq$. This implies $ac(q) = c(q)$. We can write $q = c(q) \cdot q_1, g = c(g) \cdot g_1$, with $q_1, g_1$ primitive in $R[X]$. Thus $ac(g) \cdot g_1 = fc(q) \cdot q_1$; simplifying by $c(q)$, we have $g_1 = fq_1$, so $f \mid g$ in $R[X]$.

Next, we must show that any nonzero non-invertible polynomial $f \in R[X]$ is a product of irreducibles. We prove this by induction on the degree of $f$. If $\deg f = 0$ and $f$ is not a unit in $R[X]$, then $f \in R^\circ$ and it has a decomposition in irreducible factors in $R$, factors that are also irreducible in $R[X]$. If $\deg f > 0$, write $f = c(f) \cdot f_1$, with $f_1$ primitive. It is sufficient to prove the existence of a decomposition for $f_1$. If $f_1$ is irreducible, we are finished; if not, $f_1$ has a proper divisor in $R[X]$, which must be a polynomial of degree strictly less than $\deg f$ ($f_1$ has no proper divisors in $R$, being primitive). Thus, $f_1 = gh$, with $g, h \in R[X]$, having degrees smaller than $f$. Applying the induction for $g$ and $h$, we infer that $f_1$ is a product of irreducibles in $R[X]$. \hfill \Box

Thus, the following rings are unique factorization domains:

$$\mathbb{Z}[X], \mathbb{Z}[X_1, \ldots, X_n], K[X_1, \ldots, X_n],$$

where $K$ is a field.

An analogous result holds for Noetherian rings: if $R$ is a commutative Noetherian ring, then $R[X]$ is Noetherian. (David Hilbert's Basis-satz).

---

### I.6 Polynomial ring arithmetic

In this section, $R$ designates a domain and $K$ its field of quotients. Recall that $R^\circ$ denotes the set of nonzero and non-invertible elements in $R$. 
The problem of deciding the irreducibility of a polynomial in $R[X]$ is important and often nontrivial. A rich collection of irreducibility criteria is thus very useful in such problems.

6.1 Remark. Given $f \in R[X]$, the following simple facts are worth remembering:

- if $\deg f = 0$, then $f \in R$. In this case, $f$ is irreducible in $R[X]$ if and only if it is irreducible in $R$. If $R = K$ ($R$ is a field), then $f$ is invertible and thus reducible.

- if $\deg f = n > 0$, then $f$ is irreducible in $R[X]$ if and only if $f$ has no non-invertible divisors of degree 0 and there are no decompositions $f = gh$, with $g, h \in R[X]$ and $1 \leq \deg g, \deg h < n$.

The fact that $f$ has no non-invertible divisors of degree 0 amounts to saying that the GCD of the coefficients of $f$ exists and is 1. In practice, when $R$ is a UFD, this condition reads “$f$ is primitive”. Recall that, in this case, $f$ is irreducible in $R[X] \iff f$ is primitive and $f$ is irreducible in $K[X]$.

If $R$ is a domain and not a field, $R[X]$ is not principal (and certainly not Euclidian), so the theorem of division with remainder does not hold in $R[X]$. However, if $f, g \in R[X]$, and $g$ has as leading coefficient a unit, the argument of the proof of the division with remainder theorem for $K[X]$ ($K$ a field) still holds (see Example 2.5.b). The proof of the following result is left to the reader:

6.2 Proposition. (Integer division theorem) Let $f, g \in R[X]$. If the leading coefficient of $g$ is a unit in $R$, then there exist $q, r \in R[X]$ such that $f = gq + r$, with $r = 0$ or $\deg r < \deg f$.

An important consequence is the “Theorem of Bézout”:

6.3 Theorem. Let $f \in R[X]$ and $a \in R$. The following statements are equivalent:
a) a is a root of f.

b) the polynomial $X - a$ divides f in $R[X]$.

Proof. There exist $q, r \in R[X]$ such that $f = (X - a)q + r$, where $\deg r = 0$ or $r = 0$. Note that $X - a$ divides f if and only if $r = 0$. But $f(a) = (a - a)q(a) + r(a) = r$, so $f(a) = 0$ is equivalent to $r = 0$.

This theorem is the basis of the notion of multiple root:

6.4 Definition. If $f \in R[X]$ and $a \in R$, $a$ is called a multiple root of multiplicity $m$ for $f$ if $(X - a)^m | f$ and $(X - a)^{m+1} \nmid f$; the natural number $m$ is called the multiplicity of the root $a$. A root of multiplicity 1 is called a simple root.

6.5 Corollary. Let $f \in R[X]$, $\deg f > 1$. If $f$ has a root $a \in R$, then $f$ is reducible in $R[X]$ (it is divisible with $X - a$).

The converse of this statement is false: $(X^2 + 1)^2$ has no roots in $\mathbb{Q}$, yet it is obviously reducible in $\mathbb{Q}[X]$. Nevertheless, we have:

6.6 Proposition. Let $K$ be a field. Then a polynomial $f$ of degree 2 or 3 in $K[X]$ is irreducible if and only if $f$ has no roots in $K$. In particular, if $R$ is a UFD, a primitive polynomial of degree 2 or 3 in $R[X]$ is irreducible in $R[X]$ if and only if it has no roots in $K$.

Proof. Since $f$ is irreducible in $K[X]$ and $\deg f > 1$, $f$ has no roots in $K$. Conversely, if $f$ is reducible and has degree 2 or 3, then, by looking at the degrees of the factors in a decomposition of $f$, one concludes that $f$ has a divisor of degree 1, which has a root in $K$. The remaining claims follow from “$f$ is irreducible in $R[X]$ if and only if $f$ is primitive and irreducible in $K[X]$”.

If the ring $R$ is not a UFD, then the criterion above may not work:

6.7 Examples. a) $f = (2X + 1)^2$ is reducible in $\mathbb{Z}[X]$, but has no roots in $\mathbb{Z}$. Of course, $f$ has roots in $\mathbb{Q}$, the field of quotients of $\mathbb{Z}$.
I. Arithmetic in integral domains

b) Let $R = \{a + 2bi \mid a, b \in \mathbb{Z}\}$. A quick check shows that $R$ is a subring in $\mathbb{Z}[i]$, so it is a domain. The polynomial $X^2 + 1$ is irreducible in $R[X]$ (prove!), but has the roots $i, -i$ in $\mathbb{Q}[i]$, the field of quotients of $R$. This means that a polynomial of degree 2 or 3 in $R[X]$ that has roots in $\mathbb{Q}(R)$ is not necessarily reducible in $R[X]$.

The following criterion is widely used to find all rational roots of a polynomial in $\mathbb{Z}[X]$ (also see Exercise 26).

6.8 Proposition. Let $R$ be a UFD and $f = a_0 + a_1X + \ldots + a_nX^n \in R[X]$. If $p/q \in K$ is a root of $f$, with $p, q \in R$, $(p, q) = 1$, then $p \mid a_0$ and $q \mid a_n$.

Proof. We remark first that every element of $K$ can be written as $p/q$, with $p$ and $q$ coprime. Writing that $f(p/q) = 0$ and multiplying with $q^n$, we have:

$$-a_0q^n = a_1pq^{n-1} + \ldots + a_np^n,$$

so $p \mid a_0q^n$. Since $(p, q) = 1$, we also have $(p, q^n) = 1$ ($R$ is a UFD), so $p \mid a_0$. A similar proof can be given for $q \mid a_n$.

6.9 Example. Let $f = X^3 - X + 2 \in \mathbb{Z}[X]$. If $p/q \in \mathbb{Q}$ is a root of $f$, $(p, q) = 1$, then $p \mid 2$ and $q \mid 1$. So, the rational roots of $f$ belong to the set $\{1, -1, 2, -2\}$. By direct testing, we get that none of these numbers is a root. So, $f$ has no rational roots. Since $f$ has degree 3, $f$ is irreducible in $\mathbb{Q}[X]$ (also in $\mathbb{Z}[X]$, being primitive).

Here are some general tricks that may prove useful in irreducibility problems.

6.10 Proposition. Let $f = a_0 + a_1X + \ldots + a_nX^n \in R[X], f \neq 0$.

a) Let $c, d \in R$, where $c$ is a unit in $R$. Then $f$ is irreducible if and only if $f(cX + d)$ is irreducible.

b) Suppose $f(0) = a_0 \neq 0$. Then $f$ is irreducible if and only if $r(f) = a_n + a_{n-1}X + \ldots + a_0X^n$
(the reciprocal of \(f\)) is irreducible.

c) Suppose \(f\) has no divisors of degree 0 other than units. If \(S\) is a commutative ring and \(\varphi : R \to S\) is a unitary ring homomorphism such that \(\varphi(a_n) \neq 0\) and \(\varphi(a_0) + \varphi(a_1)X + \ldots + \varphi(a_n)X^n\) is an irreducible polynomial in \(S[X]\), then \(f\) is irreducible in \(R[X]\).

d) (Eisenstein criterion) Let \(R\) be a UFD. If there exists a prime element \(p \in R\) such that \(p \mid a_i, \forall \ i < n, \ p \nmid a_n, \ p^2 \nmid a_0\), then \(f\) is irreducible in \(K[X]\) (thus \(f\) is irreducible in \(R[X]\) if it is primitive).

Proof. a) Let \(\varphi : R[X] \to R[X]\) be the unique homomorphism of \(R\)-algebras (that is, \(\psi\) is a ring homomorphism and \(\psi|_R = \varphi\)) satisfying \(\varphi(X) = cX + d\). In other words, \(\varphi(f)\) is obtained by replacing the indeterminate \(X\) in \(f\) with \(cX + d\). The element \(c\) is a unit if and only if \(\varphi\) is an isomorphism of \(R\)-algebras (the homomorphism of \(R\)-algebras \(\psi : R[X] \to R[X]\) that takes \(X\) to \(c^{-1}X - c^{-1}d\) is the inverse of \(\varphi\)). Therefore, \(f = gh \iff \varphi(f) = \varphi(g)\varphi(h), \forall g, h \in R[X]\). Since \(\varphi\) preserves the degrees and \(\varphi|_R = \text{id}_R\), one obtains that: \(f\) is irreducible if and only if \(\varphi(f)\) is irreducible.

b) If \(g, h \in R[X]\), with nonzero \(g(0)\) and \(h(0)\), then \(r(gh) = r(g)r(h)\).

Indeed, note that \(r(f) = X^n f\left(\frac{1}{X}\right)\), where \(n = \deg f\) (for a rigorous argument, consider the equality in \(K(X)\), the field of quotients of \(K[X]\)). So, if \(\deg g = m, \deg h = p\), we have

\[
\begin{align*}
\text{r}(gh) &= X^{m+p}(gh)\left(\frac{1}{X}\right) = X^m g\left(\frac{1}{X}\right)X^p h\left(\frac{1}{X}\right) = r(g)r(h).
\end{align*}
\]

Because \(r\) preserves the degrees and, for any \(d \in R\), we have \(d \mid f \iff r(d) \mid r(f)\), we get the conclusion.

c) Let \(\psi : R[X] \to S[X]\) be the unique homomorphism of \(R\)-algebras such that \(\psi(X) = X\). We must prove that \(\psi(f)\) is irreducible implies \(f\) is irreducible. Suppose \(f = gh\), with \(g, h \in R[X]\). Then \(\psi(f) = \psi(g)\psi(h)\); the condition \(\varphi(a_n) \neq 0\) ensures that \(\deg \psi(g) + \deg \psi(h) = \deg \psi(f) = n\). Since \(\deg \psi(q) \leq \deg q, \forall q \in R[X]\), we obtain that \(\deg \psi(g) = \deg g\) and
\[ \deg \psi(h) = \deg h. \] But \( \psi(f) \) is irreducible, so \( \psi(g) \) (to make a choice) is a unit, of degree 0. So, \( 0 = \deg \psi(g) = \deg g. \) Since \( f \) has no 0 degree non-invertible divisors, \( g \in U(R). \)

d) Write \( f = c(f) f_1, \) with \( f_1 \) primitive. We have that \( f \) and \( f_1 \) are associated in \( K[X]. \) By replacing \( f \) with \( f_1, \) we can assume \( f \) is primitive. It is sufficient now to prove that \( f \) is irreducible in \( R[X]. \) If \( f \) were reducible, then:

\[ f = a_0 + a_1 X + \ldots + a_n X^n = (b_0 + b_1 X + \ldots + b_m X^m) (c_0 + c_1 X + \ldots + c_p X^p), \]

where \( m > 0, p > 0, b_0, b_1, \ldots, b_m, c_0, c_1, \ldots, c_p \in R, b_m \neq 0, c_p \neq 0. \) We have \( b_0 c_0 = a_0, \) so \( p \mid b_0 c_0 \) and \( p^2 \nmid b_0 c_0; \) thus \( p \) divides exactly one of \( b_0 \) and \( c_0. \) Suppose \( p \mid b_0 \) and \( p \nmid c_0. \) Because \( p \nmid a_n, \) \( p \) does not divide all the \( b_i \)'s; thus there exists some \( i, 1 \leq i \leq m, \) such that \( p \nmid b_i \) and \( p \mid b_j, \) \( \forall j < i. \) Then \( p \nmid b_i c_0 \) and

\[ a_i = b_i c_0 + \sum_{j=1}^{i-1} b_j c_{i-j} \]

is not divisible by \( p, \) contradicting the hypothesis. \[ \square \]

**6.11 Remark.** If \( f \in R[X] \) is a **monic reducible** polynomial, there exists a decomposition of \( f \) of the form \( f = gh, \) where \( g, h \in R[X] \) are **monic**, of degrees \( > 1. \) The proof is proposed as an exercise. This simple remark is useful in reducibility issues.

A few instances of using the above criteria on, concrete cases will give an idea on the strategies of approaching the problem of irreducibility of a polynomial. The exist **algorithms** that decide if a given polynomial in \( \mathbb{Z}[X] \) is irreducible. Such an algorithm (due to Kronecker) that also outputs a factor of the polynomial if it is not irreducible is described in Exercise 31. This algorithm, applied repeatedly, yields a **factorization algorithm** (producing a decomposition in irreducible factors) for any polynomial in \( \mathbb{Z}[X] \) or \( \mathbb{Q}[X]. \) The modern symbolic computation software (**Maple, Mathematica, Macaulay, Axiom**, etc.) have powerful routines that decide polynomial irreducibility issues.
I.6 Polynomial ring arithmetic

bility, including polynomials in several indeterminates, polynomials with coefficients in algebraic extensions of \( \mathbb{Q} \) or in a finite field. One can prove that, if there exists a factorization algorithm for \( K[X] \), with \( K \) a field, then there exists one or \( L[X] \), where \( L \) is a finitely generated extension of \( K \). For details and developments, see Spindler [1994], Winkler[1996].

6.12 Examples. a) The polynomial \( 6X^9 + 13X^2 + 26 \) is irreducible in \( \mathbb{Q}[X] \) (and in \( \mathbb{Z}[X] \), it is primitive), by the Eisenstein criterion with \( p = 13 \).

b) For any prime \( p \) and any \( n \in \mathbb{N}^* \), \( X^n - p \) is irreducible in \( \mathbb{Q}[X] \) and in \( \mathbb{Z}[X] \) (use Eisenstein again).

c) Let \( p \) be a prime number and let
\[
f = X^{p-1} + X^{p-2} + \ldots + X + 1 \in \mathbb{Z}[X].
\]
The Eisenstein criterion cannot be applied directly to \( f \). Consider the polynomial
\[
g = f(X+1) = \frac{(X+1)^p - 1}{X+1 - 1} = \sum_{i=1}^{p} \binom{p}{i} X^{i-1}
\]
Remark that the Eisenstein criterion can be applied to \( g \), since \( p \) divides all the binomial coefficients \( \binom{p}{i} \), with \( 1 \leq i < p \). Thus, \( g \) is irreducible and, by a) above, \( f \) is irreducible.

d) The polynomial \( f = Y^9 + X^9 Y^7 - 3X^2 Y + 2X \) is irreducible in \( \mathbb{Z}[X, Y] \). Indeed, consider \( f \) as a polynomial in \( Y \) with coefficients in \( \mathbb{Z}[X] \), a UFD. Apply Eisenstein with \( p = X \) (\( X \) is irreducible in \( \mathbb{Z}[X] \)). Note that \( \mathbb{Z} \) can be replaced with any UFD of characteristic \( \neq 2 \).
e) Consider \( f = X^5 + X^2 + 1 \in \mathbb{Z}_2[X] \). The polynomial \( f \) has no roots in \( \mathbb{Z}_2 \) (easy check: \( \mathbb{Z}_2 \) has 2 elements, 0 and 1), so the proper divisors of \( f \) can be of degrees 2 or 3. A decomposition of \( f \) may look only like:
\[
X^5 + X^2 + 1 = (X^3 + aX^2 + bX + 1)(X^2 + cX + 1),
\]
with \( a, b, c \in \mathbb{Z}_2 \). Identifying the coefficients, we obtain a system of equations in \( a, b, c \), which is readily seen to have no solutions in \( \mathbb{Z}_2 \). Hence, \( f \) is irreducible in \( \mathbb{Z}_2[X] \).

f) A typical use of 6.10.c) for a polynomial \( f \) with integer coefficients is to “reduce the coefficients modulo \( n \)”. More precisely, for some \( n \in \mathbb{N} \) conveniently chosen, consider the unique ring homomorphism \( \phi: \mathbb{Z} \rightarrow \mathbb{Z}_n \) and investigate the “polynomial \( f \) reduced modulo \( n \)”, (denoted with \( \psi(f) \) in the Proof). Take, for instance, \( f = 7X^5 + 4X^3 - X^2 + 6X + 9 \in \mathbb{Z}[X] \). The polynomial \( f \) reduced modulo 2 is \( X^5 + X^2 + 1 \in \mathbb{Z}_2[X] \), which is irreducible. The conditions in 6.10.c) are satisfied, so \( f \) is irreducible in \( \mathbb{Z}[X] \) (also in \( \mathbb{Q}[X] \)).

\( g) 10X^7 + 5X^2 + 2 \) is irreducible in \( \mathbb{Z}[X] \): its reciprocal is \( 2X^7 + 5X^5 + 10 \), irreducible by Eisenstein with \( p = 5 \).

Exercises

Throughout the exercises, \( R \) is a domain and \( K \) is its field of quotients (unless specified otherwise).

1. Let \( R \) be a commutative unitary ring that has zero divisors and let \( a \in R \) be a zero divisor. If \( b \in R \setminus \{0\} \) is not a zero divisor, prove that \( ax + b = 0 \) has no solutions in any ring \( S \) that includes \( R \) as a subring.

2. Let \( R \) be a finite domain. Prove that \( R \) is a field.

3. Let \( R \) be an infinite unitary ring. Prove that the set \( R^\circ \) of the non-zero non-invertible elements in \( R \) is infinite. (\textit{Hint:} If \( R^\circ \) is finite, then \( U(R) \) is infinite. Let \( S(R^\circ) \) be the set of all bijections from \( R^\circ \) to \( R^\circ \). The mapping \( \varphi: U(R) \rightarrow S(R^\circ), x \mapsto \varphi_x, \varphi_x(a) = xa, \forall a \in R^\circ, \) is injective, contradiction.)
4. Let $R$ be a GCD domain. Then any element in $K$ can be written as $a/b$, $(b \neq 0)$, with $a, b \in R$, coprime. What can you say about the uniqueness of such a representation?

5. Show that a commutative ring $R$ is a domain if and only if $R[X]$ is a domain.

6. Let $p \in R^\circ$. Prove that the ideal generated by $p$ in $R[X]$, $pR[X]$, is prime if and only if $p$ is a prime element in $R$. (Hint: $R[X]/(pR[X]) \cong (R/pR)[X]$). Deduce a new proof for 5.11.a).

7. Let $d \in \mathbb{Z}$, $d < 0$ be squarefree. Determine $U(\mathbb{Z}[\sqrt{d}])$.

8. The elements 6 and $2 + \sqrt{-5}$ have no GCD (and thus no LCM) in $\mathbb{Z}[\sqrt{-5}]$.

9. Show that $\mathbb{Z}[\sqrt{2}]$ is Euclidian. (Hint. Use 3.4.)

10. Let $\theta = (1 + \sqrt{5})/2$. Show that the norm $N : \mathbb{Z}[\theta] \rightarrow \mathbb{Z}$ is $N(a + b\theta) = a^2 + ab - b^2$, $\forall a, b \in \mathbb{Z}$. Prove that $\mathbb{Z}[(1 + \sqrt{5})/2]$ is Euclidian. (Hint. Use 3.4 and $\mathbb{Q}[(\sqrt{5})] = \mathbb{Q}[\theta]$).

11. Let $\theta = (1 + i\sqrt{3})/2$. Write a formula for the norm $N : \mathbb{Z}[\theta] \rightarrow \mathbb{Z}$ and determine $U(\mathbb{Z}[\theta])$.

12. Let $d \in \mathbb{Z}$ be squarefree.

   a) Any element in $\mathbb{Q}[(\sqrt{d})]$ can be uniquely written as $a + b\sqrt{d}$, with $a, b \in \mathbb{Q}$.

   b) Suppose known that any quadratic integer is a root of a monic polynomial of degree 2 with integer coefficients. Show that: $x = a + b\sqrt{d} \in \mathbb{Q}[(\sqrt{d})]$ $(a, b \in \mathbb{Q})$ is a quadratic integer $\iff$ $\text{Tr}(x) = 2a \in \mathbb{Z}$ and $N(x) = a^2 - db^2 \in \mathbb{Z}$.

   c) Show that $x = a + b\sqrt{d}$ (with $a, b \in \mathbb{Q}$) is a quadratic integer $\iff$ $2a \in \mathbb{Z}$, $2b \in \mathbb{Z}$ and $4a^2 - 4b^2d \equiv 0 \pmod{4}$.

   d) Show that $R := \{\alpha \in \mathbb{Q}[(\sqrt{d})] \mid \alpha$ is a quadratic integer$\}$ is a subring of $\mathbb{Q}[(\sqrt{d})]$ and $R = \mathbb{Z}[\theta]$, where $\theta$ is given by Prop. 3.3.

   e) For $d < 0$, determine explicitly $U(R)$. 

13. Let $d \in \mathbb{Z}$ be squarefree and $\alpha, \beta \in \mathbb{Z}[\sqrt{d}]$ such that $\alpha \beta \in \mathbb{Z}$. Show that there is some $\gamma \in \mathbb{Z}[\sqrt{d}]$ and $a, b \in \mathbb{Z}$ such that $\alpha = a\gamma$ and $\beta = b\gamma$ ($\gamma$ is the conjugate of $\gamma$).

14. The purpose of the exercise is to determine all primes in $\mathbb{Z}[i]$. As a bonus, we find the primes in $\mathbb{Z}$ that can be written as a sum of two squares.

   a) Prove: for any $d \in \mathbb{Z}$ and any $a + bi \in \mathbb{Z}[i]$, $d | a + bi$ in $\mathbb{Z}[i] \iff d | a$ and $d | b$ in $\mathbb{Z}$.

   b) Suppose that $p \in \mathbb{Z}$ is a prime in $\mathbb{Z}$ and in $\mathbb{Z}[i]$. Show that the equation $x^2 + 1 = 0$ has no solutions in $\mathbb{Z}_p$ ($= \mathbb{Z}/p\mathbb{Z}$, the field of integers mod $p$).

   c) Let $p$ be prime in $\mathbb{Z}$. Then: ($\exists$) $a, b \in \mathbb{Z}$ such that $p = a^2 + b^2$ (p can be written as a sum of two squares) $\iff$ the equation $x^2 + 1 = 0$ has solutions in $\mathbb{Z}_p$.

   d) If $p$ is a prime in $\mathbb{Z}$ and $p \equiv 3(\text{mod } 4)$, then $p$ cannot be written as a sum of two squares.

   e) If $p$ is a prime in $\mathbb{Z}$ and $p \equiv 1(\text{mod } 4)$, then $\left(\frac{p-1}{2}\right)! \equiv -1(\text{mod } p)$ (Hint. Use Wilson's Theorem: $(p-1)! \equiv -1(\text{mod } p)$).

   f) Prove that a prime $p \in \mathbb{Z}$ is also prime in $\mathbb{Z}[i]$ if and only if $p \equiv 3(\text{mod } 4)$.

   g) Prove that all the prime elements in $\mathbb{Z}[i]$ (up to association in divisibility) are: $1 + i$, $1 - i$ and the primes $p$ in $\mathbb{Z}$ with $p \equiv 3(\text{mod } 4)$.

   h) Prove that a prime $p \in \mathbb{Z}$ can be written as a sum of two squares $\iff p \equiv 1(\text{mod } 4)$.

15. Let $R$ be a unitary subring of the commutative ring $S$. An element of $S$ is called integral over $R$ if it is a root of a nonzero monic polynomial in $R[X]$. Prove that, if $R$ is a GCD-domain, $K$ is its field of quotients and $x \in K$ is integral over $R$, then $x \in R$. (A domain $R$ with this property is called integrally closed).
16. Let $R$ be a PID and let $S$ be multiplicatively closed system in $R$. Then the ring of quotients $S^{-1}R$ is a PID.

17. Let $R$ be an Euclidian domain with respect to the function $\varphi$ and let $S$ be a multiplicatively closed system in $R$. Then $S^{-1}R$ is an Euclidian domain. (Hint. Take $S$ saturated. Use exercise 4.)

18. Let $R$ be a UFD and let $S$ be a multiplicatively closed system in $R$. Then $S^{-1}R$ is a UFD.

19. Does the property of a domain $R$ of being Euclidian (respectively a PID, a UFD) is inherited by the unitary subrings of $R$?

20. Let $R$ be Euclidian with respect to $\varphi$. Show that there exists $u \in R$ nonzero and not a unit with the property: $\forall x \in R, \exists q \in R$ such that $x - qu$ is a unit or 0. Find such a $u$ for $R = \mathbb{Z}$, $K[X]$. (Ind. min \{ $\varphi(v) \mid v \in R^{\circ}$\} = $\varphi(u)$ for some $u \in R^{\circ}$.)

21. Let $d \leq -13$, $d$ squarefree and denote by $R$ the ring of integers of $\mathbb{Q}[\sqrt{d}]$ (so, $R = \mathbb{Z}[\theta]$, with $\theta$ as in Prop. 3.3). Show that $U(R) = \{ -1, 1 \}$. Show that $R$ is not Euclidian. Is this result a particular case of Proposition 3.5? (Ind. If $R$ is Euclidian, let $u \in R$ given by the preceding exercise. Then, $\forall x \in R$, we have $u \mid x$ or $u \mid x \pm 1$. Take $x = 2$ and deduce that $u \in \mathbb{Z}$ and $u = \pm 2$ or $\pm 3$. Then find $y \in R$ such that $u$ does not divide any of $y$ or $y \pm 1$.)

22. Let $d \in \mathbb{Z}$ be squarefree, $d \equiv 1 \pmod{4}$. Then $\mathbb{Z}[\sqrt{d}]$ is not a GCD-domain. (Hint. 2 is irreducible and not a prime.)

23. Show that $\mathbb{Z}$ contains an infinity of prime elements not associated in divisibility to each other.

24. Let $R$ be a UFD that is not a field, such that the group of units $U(R)$ is finite. Then $R$ contains an infinity of prime elements not associated in divisibility to each other. (Hint. If $p_1, \ldots, p_n$ are all the primes – up to association in divisibility –, then there exists $m \geq 1$ such that $1 + (p_1\ldots p_n)^m \in R^{\circ}$.)
25. Let \( R \) be a UFD and let \( p \in R \) be a prime. Using the canonical homomorphism \( \pi : R \to R/pR \) and its extension to a homomorphism \( \psi : R[X] \to (R/pR)[X] \), give a new proof for the Eisenstein criterion. (Hint. If \( f = a_0 + a_1 X + \ldots + a_n X^n \) satisfies the hypothesis of the criterion and \( f = gh \), then \( \psi(f) = \pi(a_n)X^n = \psi(g)\psi(h) \). If \( \deg g, \deg h \geq 1 \), then \( g(0) \) and \( h(0) \) are divisible by \( p \).)

26. Let \( R \) be a UFD and \( f = a_0 + a_1 X + \ldots + a_n X^n \in R[X] \).
   a) If \( p/q \in K \) is a root of \( f \), where \( p, q \in R \) and \( (p, q) = 1 \), then \( p \mid a_0, q \mid a_n \) and \( (p - qr) \mid f(r), \forall r \in R \). Write down explicitly the conclusions for \( a_n = 1 \).
   b) Let \( g = a_n^{n-1}a_0 + a_n^{n-2}a_1 X + \ldots + a_n X^{n-1} + X^n \). Then \( a_n^{n-1}f(X) = g(a_n X) \). What connection is between the roots of \( g \) and the roots of \( f \)?
   c) Find the rational roots of \( 2X^3 + 5X^2 + 9X - 15 \) and \( 4X^3 - 7X^2 - 7X + 15 \).

27. Let \( K \) be a field. Prove that any nonzero polynomial \( f \in K[X] \) has at most \( \deg f \) roots in \( K \) (every root is counted with its multiplicity).

28. Let \( R \) be a commutative ring. Prove that the following statements are equivalent:
   a) Any nonzero polynomial \( f \in R[X] \) has at most \( \deg f \) roots in \( R \).
   b) Any polynomial of degree 1 has at most one root in \( R \).
   c) \( R \) is a domain.
   (Hint. Consider the field of quotients of \( R \) and use the previous problem).

29. Let \( R \) be a commutative ring. If \( f \in R[X] \), define the polynomial function \( \tilde{f} : R \to R \), defined by: \( \forall x \in R \), \( \tilde{f}(x) = f(x) \) (the value of \( f \) in \( x \)). Prove that, if \( R \) is an infinite domain, then the mapping \( \varphi : R[X] \to R^R \), \( \varphi(f) = \tilde{f} \), \( \forall f \in R[X] \), is injective. Is the conclusion still valid if one does not assume that \( R \) is infinite?

30. (The Lagrange interpolation polynomial) Let \( K \) be a field, \( n \geq 1 \) an integer, fix \( n + 1 \) distinct elements \( x_0, \ldots, x_n \in K \) and (not necessarily
distinct) $y_0, \ldots, y_n \in K$. Prove that there exists a unique polynomial $L \in K[X]$ satisfying: deg $L \leq n$ and $L(x_i) = y_i$, $0 \leq i \leq n$.

31. Let $p \in \mathbb{Z}[X]$, primitive, deg $p = n$ and let $m = \text{the largest integer} \leq n/2$.

a) Show that $p$ is reducible in $\mathbb{Z}[X] \iff p$ has a divisor of degree between 1 and $m$.

b) Choose $m + 1$ distinct integers, $(x_0, \ldots, x_m) \in \mathbb{Z}^{m+1}$. Show that the following algorithm terminates in a finite number of steps and outputs a nontrivial factor of $p$ of degree $\leq m$ or proves that $p$ is irreducible:

1. If $\exists i$ with $p(x_i) = 0$, then $X - x_i$ is a factor of $p$ and STOP. If not, go to 2.
2. Let $D = \{d = (d_0, \ldots, d_m) \in \mathbb{Z}^{m+1} \mid d_i \mid p(x_i), \forall i\}$. $D$ is a finite set.
   For any $d \in D$, let $L_d \in \mathbb{Q}[X]$ be the Lagrange interpolation polynomial with $L_d(x_i) = d_i$, $\forall i$, and deg $L_d \leq m$. If there exists $d \in D$ with $L_d \in \mathbb{Z}[X]$ and $L_d \mid p$, then $L_d$ is a factor of $p$ and STOP. If not, then $p$ is irreducible.

c) Deduce an algorithm of deciding the irreducibility of polynomials in $\mathbb{Q}[X]$.

d) Suppose $m = 2$. Propose a choice for $(x_0, \ldots, x_m)$.

e) Suppose there is available a factorization algorithm for $R$ (an algorithm that produces a decomposition of any element in $R$ in prime factors). What properties should $R$ have in order to adapt the algorithm above to $R[X]$?

f) Suppose $R$ is a UFD and there exists a factorization algorithm for $R[X]$. Then there exists a factorization algorithm for $K[X]$.

32. Decide the irreducibility of $X^4 + X^2 + 2X - 1 \in \mathbb{Z}[X]$.

33. Show that $a_0 + a_1 X + \ldots + a_n X^n \in \mathbb{Z}_2[X]$ ($a_n \neq 0$) has no roots in $\mathbb{Z}_2$ if and only if $a_0(a_0 + a_1 + \ldots + a_n) \neq 0$. 
34. Using the equality (in $\mathbb{Z}_2[X]$):
$$X^5 + X + 1 = (X^3 + X^2 + 1)(X^2 + X + 1),$$
prove that $X^5 - X - 1 \in \mathbb{Q}[X]$ is irreducible.

35. a) Let $f \in R[X]$, $\text{deg } f = m$. If $f$ has at least $m + 1$ roots in $R$, then $f = 0$.

   b) Let $g \in R[X_1, \ldots, X_n]$, with $R$ infinite. If $g(a_1, \ldots, a_n) = 0$, $\forall (a_1, \ldots, a_n) \in R^n$, then $g = 0$. Deduce that two polynomials in $R[X_1, \ldots, X_n]$ are equal if and only if the associated polynomial functions are equal.

   c) Give an example of a finite field $K$ and distinct polynomials in $K[X]$ that have the same associated polynomial function.

   d) (Generalization of b) Let $R$ be infinite and let $g \in R[X_1, \ldots, X_n]$, with $\text{deg}(g, X_i) = m_i$. Suppose that $(\exists) S \subseteq R$ with $|S| > m_1$; $(\forall) a_1 \in S$, $(\exists) S(a_1) \subseteq R$ with $|S(a_1)| > m_2$; $(\forall) a_1 \in S$, $(\forall) a_2 \in S(a_1)$, $(\exists) S(a_1, a_2) \subseteq R$ with $|S(a_1, a_2)| > m_3$ and so on. If $g(a_1, \ldots, a_n) = 0$, $\forall a_1 \in S$, $\forall a_2 \in S(a_1)$, $\forall a_3 \in S(a_1, a_2)$, $\ldots$, $\forall a_n \in S(a_1, \ldots, a_{n-1})$, then $g = 0$.

36. Let $R$ be a domain. A polynomial in $n$ indeterminates $p \in R[X_1, \ldots, X_n] =: R[X]$ is called homogeneous of degree $q$ if all the monomials in $p$ have total degree $q$ (see the Appendix). Show that:

   a) Any $p \in R[X]$ can be written uniquely as:
   $$p = p_0 + p_1 + \ldots + p_m,$$
   with $p_i \in R[X]$, homogeneous of degree $i$.

   b) The product of two homogeneous polynomials of degrees $a$, respectively $b$, is homogeneous of degree $a + b$.

   c) $p$ is homogeneous of degree $q$ $\iff$ $p(TX_1, \ldots, TX_n) = T^q p(X_1, \ldots, X_n)$ (equality in $R[X_1, \ldots, X_n, T] = R[X_1, \ldots, X_n][T]$).

   d) Let $p \in R[X]$ be homogeneous. Then any divisor of $p$ in $R[X]$ is homogeneous. (Hint. For $p = gh$, write $g = g_0 + \ldots + g_m$, where $g_i$ is homogeneous of degree $i$, $0 \leq a \leq m$ and $g_a \neq 0 \neq g_m$. It suffices to prove that $a = m$.)
e) If $R$ is infinite and $p \in R[\bar{X}]$, then $p$ is homogeneous of degree $q$ \iff $p(tx_1, \ldots, tx_n) = t^q p(x_1, \ldots, x_n), \ \forall t, x_1, \ldots, x_n \in R$.

f) Is it true that any symmetric polynomial in $R[\bar{X}]$ has all its divisors symmetric polynomials in $R[\bar{X}]$?

37. Let $K$ be a field of characteristic not equal to 2 ($1 + 1 \neq 0$ in $K$) and $p$ a homogeneous polynomial of degree 2 in $K[X, Y]$, i.e. $p = aX^2 + bXY + cY^2$, with $a, b, c \in K$. Prove that $p$ is reducible in $K[X, Y] \iff b^2 - 4ac$ is a square in $K \iff b^2 - 4ac = 0$ or there exist $\alpha, \beta \in K$, $(\alpha, \beta) \neq (0, 0)$, with $p(\alpha, \beta) = 0$.

38. Assume $K$ is a field and $p = a_0 Y^n + a_1 Y^{n-1}X + \ldots + a_n X^n$ is a homogeneous polynomial of degree $n$ in $K[X, Y]$. Let $p(X, 1) = a_0 + a_1 X + \ldots + a_n X^n \in K[X]$. Prove that, for any $g \in K[X, Y]$, $g | p$ in $K[X, Y]$ if and only if $g$ is homogeneous and $g(X, 1) | p(X, 1)$ in $K[X]$.

39. Let $K$ be a field and let $p \in K[X, Y]$ be homogeneous. Prove that $p$ is irreducible in $K[X, Y] \iff p(X, 1)$ is irreducible in $K[X]$.

40. Write a decomposition in irreducible factors for $X_1^3 + X_2^3 \in K[X_1, X_2]$. (Hint. The cases char $K = 3$ and char $K \neq 3$ should be treated separately).

41. Let $K$ be a field, char $K \neq 3$ and $f = X_1^3 + \ldots + X_n^3 \in K[X_1, \ldots, X_n]$. Show that $f$ is irreducible if and only if $n \geq 3$. Generalization. (Hint. For $n = 3$, apply Eisenstein to $f \in K[X_1, X_2][X_3]$. Use then an induction on $n$.)

42. Consider $n^2$ indeterminates $X_{ij}$, $1 \leq i, j \leq n$, and consider the $n \times n$ matrix $A = (X_{ij})_{1 \leq i, j \leq n} \in M_n(\mathbb{Z}[X_{ij}; 1 \leq i, j \leq n])$. Then the polynomial:

$$\det A = \sum \{X_{1\sigma(1)} \ldots X_{n\sigma(n)} \mid \sigma \in S_n\}$$

is irreducible in $\mathbb{Z}[X_{ij}; 1 \leq i, j \leq n]$.

43. Prove that a commutative ring $R$ is Noetherian if and only if every ideal in $R$ is finitely generated.
II. Modules

Module theory can be seen as a generalization of the classic linear algebra (which studies vector spaces over an arbitrary field\(^1\)). The theory is fundamental in many areas of mathematics: commutative algebra, algebraic number theory, group representation theory, algebra structure theorems, homological algebra etc. Also, module theory illustrates and uses concepts of category theory (we use some elementary notions from category theory in this chapter, notions that can be found in the Appendix). Module theory language and results are indispensable throughout most of modern algebra.

II.1 Modules, submodules, homomorphisms

The notion of a module over a ring can be obtained by replacing the word “field” with the word “ring” in the definition of the vector space.

\(^1\) Although the volume *Algèbre linéaire* (1961) of the famous Bourbaki series “Eléments de Mathématique” begins with the definition of the… module.
1.1 Definition. Let $R$ be a ring with identity (not necessarily commutative) and $(M, +)$ an Abelian group. We say that $M$ is a **left $R$-module** (or **left module over $R$**) if there exists an “external operation on $M$ with operators in $R$”\(^2\), i.e. a function

$$\mu : R \times M \to M$$

(notation: $\mu (r, x) =: rx$, $\forall r \in R$, $\forall x \in M$), satisfying, for any $r, s \in R$ and $x, y \in M$:

i) $r(x + y) = rx + ry$;

ii) $(r + s)x = rx + sx$;

iii) $(rs)x = r(sx)$;

iv) $1x = x$,

1.2 Remark. The addition in $R$ is denoted by $+$, as the addition in $M$. Also, the zero element in $R$ is denoted by 0, like the zero element in $M$. For instance, in axiom ii), the $+$ in the LHS denotes the addition in $R$, whereas the $+$ in the RHS denotes the addition in $R$. This notational abuse (which is widely used) should not confuse the reader.

If the axiom iii) is replaced by:

iii') $(rs)x = s(rx)$, $\forall r, s \in R$, $\forall x \in M$,

we say that $M$ is a **right $R$-module**. The usual notation for the “scalar multiplication” in the case of right $R$-modules is “with the scalars on the right”, i.e. the scalar multiplication is a function $\mu : M \times R \to M$, with the notation $\mu (x, r) = xr$, $\forall r \in R$, $\forall x \in M$. The axioms for the right $R$-module become in this case:

i') $(x + y)r = xr + yr$;

ii') $x(r + s) = xr + xs$;

iii') $x(rs) = (xr)s$;

iv') $x1 = x$,

---

\(^2\) Also called “multiplication of elements in $M$ with scalars in $R$”.
for any \( r, s \in R \) and \( x, y \in M \).

There is a handy notation for the fact that \( M \) is a left \( R \)-module, namely \( _R M \).

“\( M \) is a right \( R \)-module” is denoted by \( M_R \).

If \( R \) is commutative, then the notions of left \( R \)-module and right \( R \)-module are the same (look at axiom iii').

If \( R \) is an arbitrary ring and \( M \) is a right \( R \)-module, then \( M \) becomes a left \( R^{\text{op}} \)-module, where \((R^{\text{op}}, +, \ast)\) is the opposite of the ring \( R \) \((R^{\text{op}} \) and \( R \) have the same underlying Abelian group \( R \), but the multiplication \( \ast \) in \( R^{\text{op}} \) is defined by \( r \ast s = sr, \forall r, s \in R \)).

The construction above shows that a result that holds for any ring \( R \) and any right \( R \)-module is valid also for any left \( R \)-module, and conversely. In the same way, all definitions for left modules have a natural correspondent for right modules.

1.3 Examples. a) If \( K \) is a field, a \( K \)-module is exactly a \( K \)-vector space.

b) If \( R \) is a ring with identity, \( R \) has a (canonical) structure of left \( R \)-module, denoted \( _R R \). Indeed, \((R, +)\) is an Abelian group; the “external operation” \( R \times R \to R \) is the ring multiplication: \((r, s) \mapsto rs, \forall r, s \in R \). Similarly, \( R \) is canonically a right \( R \)-module, denoted \( RR \).

c) Any Abelian group \((A, +)\) is canonically a \( \mathbb{Z} \)-module. For \( n \in \mathbb{Z} \) and \( a \in A \), \( na \) is defined as the “multiple” of \( a \) in the additive group \( A \) (if \( n \in \mathbb{N} \), \( na = a + \ldots + a \) (\( n \) terms); if \( n < 0 \), \( na = (-a) + \ldots + (-a) \) (\( n \) terms)). This is the only external operation that endows \( A \) with a \( \mathbb{Z} \)-module structure (exercise!). The theory of Abelian groups is thus a particular case of module theory.

d) Let \( R \) be a ring and let \( n \in \mathbb{N}^* \). The \( n \)-fold Cartesian product \( R^n = \{(x_1, \ldots, x_n) \mid x_i \in R, 1 \leq i \leq n\} \) becomes an \( R \)-module if the addition and the scalar multiplication are defined component-wise:
\[
(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n), \quad (r, y_1, \ldots, y_n) \in R^n.
\]
II.1 Modules, submodules, homomorphisms

\[ r(x_1, \ldots, x_n) = (rx_1, \ldots, rx_n), \forall r \in R, \forall (x_1, \ldots, x_n) \in R^n. \]

e) If \( R \) is a ring and \( m, n \in \mathbb{N}^* \), the set \( M_{m,n}(R) \) of \( m \times n \) matrices with entries in \( R \) is an Abelian group endowed with usual matrix multiplication and becomes an \( R \)-module by defining the “multiplication of matrices with scalars”: for \( r \in R, A = (a_{ij}) \in M_{m,n}(R), \ r(a_{ij}) := (ra_{ij}) \) (multiply every entry of the matrix with \( r \)).

f) Let \( R := M_2(\mathbb{Z}) \) (the ring of 2\( \times \)2 matrices with entries in \( \mathbb{Z} \)) and \( M := M_{2,1}(\mathbb{Z}) \) (the Abelian group of 2\( \times \)1 matrices with entries in \( \mathbb{Z} \)). \( M \) has a natural structure of left \( R \)-module: \( \forall A \in M_2(\mathbb{Z}) = R, \ \forall U \in M_{2,1}(\mathbb{Z}) = M, AU \in M \) is the usual matrix product. Checking the module axioms is straightforward and it boils down to the known properties of matrix operations. Can you generalize this example? Can \( M \) be endowed with a “natural” structure of a right \( R \)-module?

g) Let \( \varphi : R \rightarrow S \) be a unitary ring homomorphism. If \( M \) is a left \( S \)-module, then \( M \) has a structure of left \( R \)-module by “\textit{restriction of scalars}”: \( \forall r \in R, \ \forall x \in M, rx := \varphi(r)x \). In particular, \( S \) becomes a left \( R \)-module (and also a right \( R \)-module). This example generalizes a situation often encountered in field extensions: any field \( S \) is a vector space over any subfield \( R \).

1.4 Remark. For a ring \( R \) and an Abelian group \( M \), defining a left \( R \)-module structure on \( M \) amounts to defining a unitary ring homomorphism \( \lambda : R \rightarrow \text{End}(M) \), where \( (\text{End}(M), +, \circ) \) is the \textit{endomorphism ring} of the Abelian group \( M \), defined as follows: “\( + \)” is homomorphism addition: \( \forall u, v \in \text{End}(M), (u + v)(x) := u(x) + v(x), \forall x \in M \), and “\( \circ \)” is the usual map composition: \( \forall u, v \in \text{End}(M), (u \circ v)(x) = u(v(x)), \forall x \in M. ^3 \)

---

\(^3\) Sometimes this ring is called the ring of \textit{left} endomorphisms of \( M \), emphasizing that the functions are written at the left of the argument, like \( u(x) \); this forces the definition of the composition of functions in the “usual” manner defined above. But...
Indeed, if $M$ is a left $R$-module, define $\lambda : R \to \text{End}(M)$ by $\lambda(r)(x) = rx$, $\forall r \in R$, $\forall x \in M$. Conversely, given $\lambda : R \to \text{End}(M)$, define scalar multiplication $R \times M \to M$ by $(r, x) \mapsto rx := \lambda(r)(x)$, $\forall r \in R$, $\forall x \in M$. The reader should check the details. What corresponds to a structure of right $R$-module of $M$?

Throughout this section, $R$ denotes a ring with identity. All modules are left $R$-modules (unless specified otherwise).

1.5 Proposition. Let $M$ be an $R$-module. The, for any $x \in M$ and $r \in R$, we have:

a) $0x = r0 = 0$.

b) $r(−x) = (−r)x = −(rx)$.

Proof. a) $0x = (0 + 0)x = 0x + 0x$. Since $M$ is a group, by simplification we deduce that $0x = 0$. The same method shows $r0 = 0$.

b) $0 = r0 = r(x + (−x)) = rx + r(−x)$. So $−(rx)$, the opposite of $rx$ in $(M, +)$, is $r(−x)$.

We define the natural notion of submodule:

1.6 Definition. a) Let $M$ be a left $R$-module. A non-empty subset $L$ of $M$ is called left $R$-submodule of $M$ if

i) $L$ is a subgroup in $(M, +)$: $\forall x, y \in L \Rightarrow x − y \in L$;

ii) $\forall r \in R$, $\forall x \in L \Rightarrow rx \in L$.

Usually one says $L$ is a submodule of $M$ if no confusions can occur. Notation: $L \leq_R M$ (or, simpler, $L \leq M$). The fact that $L$ is a right $R$-submodule of $M$ is written $L \leq M_R$.

if one writes $(x)u$ for the value of $u$ at $x$, then the composition of $u$ and $v$ is defined as $(x)(uv) = ((x)u)v$. With this multiplication, $\text{End}(M)$ is called the ring of right endomorphisms of $M$ and it is the opposite of the ring of left endomorphisms of $M$. 

The definition above is the natural generalization of the notion of vector subspace.

1.7 Proposition. Let $M$ be an $R$-module and $\emptyset \neq L \subseteq M$. The following statements are equivalent:

a) $L \leq_{R} M$: $(\forall r \in R, \forall x, y \in L \Rightarrow x - y \in L$ and $rx \in L)$.

b) For any $r, s \in R$ and $x, y \in L$, it follows that $rx + sy \in L$.

c) For any $n \in \mathbb{N}^{*}$, $r_{1}, \ldots, r_{n} \in R$, $x_{1}, \ldots, x_{n} \in L$, it follows that $r_{1}x_{1} + \ldots + r_{n}x_{n} \in L$.

Proof. a)$\Rightarrow$ b) If $r, s \in R$ and $x, y \in L$, then a) implies that $rx$ and $(-s)y \in L$, so $L$ contains also $rx - (-s)y = rx - (-sy) = rx + sy$.

b)$\Rightarrow$ a) If $r \in R$ and $x, y \in L$, then $rx + 0y = rx \in L$ and $1x + (-1)y = x - y \in L$.

b)$\Rightarrow$ c) Induction on $n$ (exercise).

An element of the form $r_{1}x_{1} + \ldots + r_{n}x_{n}$, with $r_{1}, \ldots, r_{n} \in R$ and $x_{1}, \ldots, x_{n} \in M$, is called a linear combination of $x_{1}, \ldots, x_{n}$ ($r_{1}, \ldots, r_{n}$ are called the coefficients of the linear combination). Thus, $L \leq_{R} M$ iff any linear combination of elements in $L$ is still in $L$.

Since any submodule $L$ of a module $M$ is a subgroup of the additive group $(M, +)$, we have $0 \in L$. Also, if $L \leq_{R} M$, $L$ has a structure of left $R$-module: the external operation is the restriction at $R \times L$ of the external operation of $M$.

1.8 Examples. a) For any $R$-module $M$, $\{0\}$ is an $R$-submodule in $M$, denoted simply by 0. Also, $M$ is an $R$-submodule of $M$. A submodule of $M$, not equal to $M$, is called a proper submodule of $M$.

b) The left submodules of the canonical module $_R R$ are exactly the left ideals of the ring $R$; the notation $I \leq_{R} R$ means “$I$ is a left ideal of $R$”.

c) If $M$ is an Abelian group (= $\mathbb{Z}$-module), a $\mathbb{Z}$-submodule of $M$ is the same thing with a subgroup of $M$. 
1.9 Proposition. Let \((M_i)_{i \in I}\) be a family of submodules of \(RM\). Then the intersection of this family, \(\bigcap_{i \in I} M_i\), is a submodule of \(M\). 

This simple result allows us to define the notion of submodule generated by a subset:

1.10 Definition. Let \(M\) be an \(R\)-module and let \(X\) be a subset of \(M\).

a) The intersection of all submodules of \(M\) that include \(X\) is a submodule of \(M\), called the submodule generated by \(X\) and denoted by \(R<X>\) (or simply \(<X>\)). If \(L \leq RM\) and \(<X> = L\), one says also that \(X\) is a system of generators for \(L\) (or that \(X\) generates \(L\)).

b) Define the set of linear combinations of elements in \(X\) with coefficients in \(R\) as the set \(RX\), where

\[ RX := \{ r_1 x_1 + ... + r_n x_n \mid n \in \mathbb{N}, r_1, ..., r_n \in R, x_1, ..., x_n \in X \}. \]

If \(X = \emptyset\), define \(R\emptyset = \{0\}\).

1.11 Proposition. Let \(M\) be an \(R\)-module and \(X \subseteq M\). Then:

a) \(<X>\) is the smallest (inclusion-wise) submodule of \(M\) that includes \(X\).

b) \(<X> = RX\), that is, the submodule generated by \(X\) is the same as the set of linear combinations of elements in \(X\) with coefficients in \(R\).

Proof. a) Evidently, \(<X>\) is a submodule and includes \(X\). If \(L\) is a submodule in \(M\) that includes \(X\), then \(<X> \subseteq L\) because \(L\) is a member of the family of submodules the intersection of which is \(L\).

b) We show that \(X \subseteq RX\) and that \(RX\) is the smallest submodule that includes \(X\). The case \(X = \emptyset\) is trivial. If \(X \neq \emptyset\) and \(x \in X\), then \(x\) is a linear combination, \(x = 1x \in RX\). So \(X \subseteq RX\). The difference of two linear combinations in \(RX\) and the product of any \(r \in R\) with a linear combination is still in \(RX\). So, \(RX\) is a submodule. If \(L\) is a submodule that includes \(X\), 1.7.c) implies \(RX \subseteq L\). 

1.12 Definition. For any \(a \in RM\), the submodule generated by \(\{a\}\) is \(Ra = \{ra \mid r \in R\}\) and it is also called the cyclic submodule
generated by $a$. The $R$-module $M$ is called finitely generated if there exists a finite system of generators for $M$, i.e. a finite subset $F$ of $M$ such that $\langle F \rangle = M$.

While the intersection of a family of submodules is a submodule, the union of a family of submodules is not a submodule in general.

1.13 Definition. Let $M$ be an $R$-module and let $E, F$ be submodules in $M$. The submodule generated by $E \cup F$ is called the sum of the submodules $E$ and $F$, and is denoted $E + F$. Thus, $E + F$ is just another notation for $\langle E \cup F \rangle$.

For an arbitrary family $(E_i)_{i \in I}$ of submodules of $M$, the submodule generated by $\bigcup_{i \in I} E_i$ is called the sum of the family of submodules $(E_i)_{i \in I}$, denoted $\sum_{i \in I} E_i$ or $\sum_i E_i$.

The sum of the submodules $E_1, \ldots, E_n$ is denoted $E_1 + \ldots + E_n$ or $\sum_{i=1}^n E_i$.

The sum of the family of submodules $(E_i)_{i \in I}$ is the smallest submodule of $M$ including all submodules $E_i$.

1.14 Proposition. a) If $E, F$ are submodules of $\mathbb{R}M$, then the sum of $E$ and $F$ is

$$E + F = \{ e + f \mid e \in E, f \in F \},$$

b) If $E_1, \ldots, E_n$ are submodules of $M$, then

$$E_1 + \ldots + E_n = \{ e_1 + \ldots + e_n \mid e_1 \in E_1, \ldots, e_n \in E_n \}.$$

Proof. a) Let $S := \{ e + f \mid e \in E, f \in F \}$. A straightforward verification shows that $S$ is a submodule. If $L$ is a submodule containing $E$ and $F$, then $e + f \in L$, $\forall e \in E, \forall f \in F$. Thus, $S \subseteq L$ and $S$ is the smallest submodule including $E \cup F$. \qed

In order to formulate a similar result for the case of the sum of an arbitrary (possibly infinite) family of submodules, we introduce the following notion: for a set $I$ (seen as a set of indexes), an $R$-module $M$
and a family of elements\(^4\) \((e_i)_{i \in I}\), with \(e_i \in M, \forall i \in I\), define the support of the family (denoted Supp((\(e_i)_{i \in I}\))):

\[
\text{Supp}((e_i)_{i \in I}) := \{i \in I \mid e_i \neq 0\}
\]

For any family \((e_i)_{i \in I}\) having finite support \(J \subseteq I\), its sum is defined as

\[
\sum_{i \in I} e_i := \sum_{i \in J} e_i.
\]

1.15 Proposition. If \((E_i)_{i \in I}\) is a family of submodules of \(M\), then

\[
\sum_{i \in I} E_i = \left\{ \sum_{i \in I} e_i \mid e_i \in E_i, \forall i \in I, \sum_{i \in I}(e_i)_{i \in I} \text{ having finite support} \right\} = \\
\{ e_{i_1} + \ldots + e_{i_n} \mid n \in \mathbb{N}, i_1, \ldots, i_n \in I, e_{i_1} \in E_{i_1}, \ldots, e_{i_n} \in E_{i_n} \}.
\]

Proof. Let \(S = \left\{ \sum_{i \in I} e_i \mid e_i \in E_i, \forall i \in I, \sum_{i \in I}(e_i)_{i \in I} \text{ having finite support} \right\}\). As above, we show that \(S\) is a submodule: if \(r \in R\), and \(e = \sum_{i \in I} e_i \in S\), with \(e_i \in E_i, \forall i \in I\), and \(\text{Supp}(e_i)_{i \in I}\) finite, then \(re = \sum_{i \in I} re_i \in S\), since \(E_i\) is a submodule. Likewise, if \(e, f \in S\), then \(e - f \in S\). On the other hand, clearly \(E_i \subseteq S, \forall i \in I\). If \(L\) is another submodule of \(M\) that includes all submodules \(E_i\), then \(S \subseteq L\). So \(S = \bigcup_{i \in I} E_i\). \(\square\)

1.16 Remark. The set \(\mathcal{L}_R(M)\) of all submodules of the \(R\)-module \(M\) is ordered by inclusion; furthermore, \((\mathcal{L}_R(M), \subseteq)\) is a complete lattice:\(^5\) for any subset \(\mathcal{F}\) of \(\mathcal{L}_R(M)\) (i.e., any family of submodules of \(M\)), \(\text{sup} \mathcal{F}\) is the sum of the family \(\mathcal{F}\), and \(\text{inf} \mathcal{F} = \bigcap \mathcal{F}\).

---

\(^4\) Note that a “family \((e_i)_{i \in I}\) of elements of \(M\)” is in fact a function \(f : I \rightarrow M\) (denoting \(f(i) = e_i, \forall i \in I\)).

\(^5\) An ordered set \(A\) is called a complete lattice if, for any \(B \subseteq A\), there exists \(\text{sup} B\) (the smallest upper bound of \(B\)) and \(\text{inf} B\) (the largest lower bound of \(B\)) in \(A\).
1.17 Remark. The sum of the submodules $I$, $J$ of the left canonical module $R R$ is the same with the sum of the left ideals $I$ and $J$. If $R$ is a field, the submodules of an $R$-module $M$ are the vector subspaces of $M$ “sum of submodules” means “sum of vector subspaces”.

The lattice $\mathcal{L}_R(M)$ always has a greatest element ($M$ itself) and a smallest element (the submodule 0). For this reason, the notions of maximal submodule and minimal submodule are defined as follows.

1.18 Definition. The submodule $L$ of the $R$-module $M$ is called a maximal submodule if $L$ is maximal among the proper (distinct from $M$) submodules, i.e.:

$$\forall E \leq_R M \text{ with } E \neq M, L \subseteq E \implies L = E.$$  

The submodule $L \leq M$ is called a minimal submodule if $L$ is minimal among the nonzero submodules:

$$\forall E \leq_R M \text{ with } E \neq 0, E \subseteq L \implies E = L.$$  

The following theorem is very important.

1.19 Theorem. Let $M$ be a nonzero finitely generated $R$-module. Then any proper submodule of $M$ is included in some maximal submodule. In particular, $M$ has a maximal submodule.

Proof. Let $L \leq_R M$, $L \neq M$ and let $\{x_1, \ldots, x_n\}$ be a finite generator set of $M$. Let $\mathbf{P}$ be the set of proper submodules of $M$ that include $L$. $\mathbf{P}$ is ordered by inclusion; its maximal elements (if they exist!) are exactly the maximal submodules of $M$ that include $L$. We use Zorn's lemma to prove that maximal elements exist in $\mathbf{P}$. First, note that $\mathbf{P} \neq \emptyset$ since $L \in \mathbf{P}$. Take a chain $(E_i)_{i \in I}$, with $E_i \in \mathbf{P}$, $\forall i \in I$. This chain of submodules is bounded above in $\mathbf{P}$ by $\bigcup_{i \in I} E_i =: E$. Indeed, $E$ is a submodule$^6$: if $x, y \in E$, then, for some $i, j \in I$, $x \in E_i$ and $y \in E_j$;  

$^6$ Here is a (singular) situation when the union of a family of submodules is a submodule.
(Ei)i∈I being a chain, we have Ei ⊆ Ej or Ej ⊆ Ei. So x − y ∈ Ei or Ej. Anyway, x − y ∈ E. Similarly, ∀r ∈ R, ∀x ∈ E, we have rx ∈ E. So, E ≤ M and E includes L.

We must also prove that E ≠ M. Suppose E = M. Then \{x_1, ..., x_n\} ⊆ E = \bigcup_{i \in I} E_i, so, ∀t ∈ \{1, ..., n\}, there exists \(i_t \in I\) such that \(x_t \in E_{i_t}\). But (Ei)i∈I is a chain, so \exists j \in \{i_1, ..., i_n\} such that \(E_{i_t} \subseteq E_j, \forall t \in \{1, ..., n\}\). Thus \(\{x_1, ..., x_n\} \subseteq E_j\). We deduce \(M = <x_1, ..., x_n> \subseteq E_j\), contradicting \(E_j \neq M\) (since \(E_j \in \mathcal{P}\)).

Zorn's Lemma provides us with a maximal element of \(\mathcal{P}\).

Taking \(L = 0\), the existence of a maximal submodule in \(M\) is proven. \(\square\)

1.20 Corollary (Krull's Lemma\(^7\)) Let \(R\) be a ring with identity. Then every left proper ideal of \(R\) is included in some maximal left ideal. In particular, \(R\) has a maximal left ideal.

Proof. The canonical left \(R\)-module \(RR\) is finitely generated (by \(\{1\}\)). \(\square\)

1.21 Definition. Let \(M\) and \(N\) be left \(R\)-modules. A function \(\varphi : M \to N\) is called a left \(R\)-module homomorphism (or simply module homomorphism) or \(R\)-homomorphism if it preserves the module operations:

\[ \varphi(x + y) = \varphi(x) + \varphi(y), \forall x, y \in M; \]
\[ \varphi(rx) = r\varphi(x), \forall x \in M, \forall r \in R. \]

Other names: \(R\)-linear application, linear transformation\(^8\), \(R\)-morphism. An \(R\)-module homomorphism \(\varphi : M \to M\) is called an endomorphism of \(M\).

\(^7\)Wolfgang Adolf Ludwig Helmuth Krull (1899-1971), German mathematician.
\(^8\)A geometric terminology, used mainly for vector space homomorphisms.
The first condition in the definition of the module homomorphism \( \varphi : M \to N \) means that \( \varphi \) is an Abelian group homomorphism. Thus,
\[
\varphi(0) = 0
\]
(0 denotes the zero element of \( M \), as well as the zero element of \( N \)).
Also:
\[
\varphi(-x) = -\varphi(x), \forall x \in M.
\]

1.22 Remark. Let \( L \subseteq M \). If \( L \) has an \( R \)-module structure such that the canonical inclusion \( \iota : L \to M \) is a module homomorphism, then \( L \leq M \). Conversely, if \( L \leq M \), then \( \iota : L \to M \) is a module homomorphism.

1.23 Examples. a) For any \( R \)-modules \( M \) and \( N \), \( 0 : M \to N \), \( 0(x) = 0, \forall x \in M \), is an \( R \)-module homomorphism, called the zero homomorphism. The identity application \( \text{id}_M : M \to M \), \( \text{id}_M(x) = x, \forall x \in M \), is also a homomorphism, the identity homomorphism of \( M \).

b) If \( M \) is an \( R \)-module and \( x \in M \), the “multiplication by \( x \)”, \( r_x : R \to M \) defined by \( r_x(a) = ax, \forall a \in R \), is a module homomorphism. Indeed,
\[
\begin{align*}
r_x(a + b) &= (a + b)x = ax + bx = r_x(a) + r_x(b); \\
r_x(ba) &= (ba)x = b(ax) = br_x(a), \forall a, b \in R.
\end{align*}
\]

c) If \( R \) is commutative ring and \( r \in R \), then the “multiplication by \( r \)”, \( \lambda_r : M \to M \), \( \lambda_r(x) = rx, \forall x \in M \), is a homomorphism: it is obviously additive and
\[
\lambda_r(ax) = r(ax) = (ra)x = (ar)x = a\lambda_r(x), \forall x \in M, \forall a \in R.
\]
Note that the commutativity of \( R \) is effectively used.

1.24 Definition. For any \( R \)-modules \( M, N \), we denote
\[
\text{Hom}_R(M, N) := \{ \varphi \mid \varphi : M \to N, \varphi \text{ is an } R \text{-module homomorphism} \};
\]
\[
\text{End}_R(M) := \text{Hom}_R(M, M).
\]
We denote sometimes:
\[
\text{Hom}_{(R} M, N), \text{ respectively } \text{End}_{(R} M) - \text{ for left modules;}
\]
Hom($M_R$, $N$), respectively End($M_R$) - for right modules.

1.25 Remark. Hom$_R(M, N)$ is always nonempty; it contains at least the zero homomorphism $0 : M \to N$, $0(x) = 0$, $\forall x \in M$. Moreover, Hom$_R(M, N)$ is an Abelian group with respect to homomorphism addition, defined below:

1.26 Proposition. a) Let $E$, $F$ be $R$-modules and let $\varphi : E \to F$, $\eta : E \to F$ be $R$-module homomorphisms. Then the sum $\varphi + \eta : E \to F$, defined by:

$$(\varphi + \eta)(x) := \varphi(x) + \eta(x), \forall x \in E,$$

is an $R$-module homomorphism. Hom$_R(E, F)$ is an Abelian group with respect to homomorphism addition; the zero element is the 0 homomorphism, the opposite of $\varphi$ is $(-\varphi)$, $(-\varphi)(x) = -\varphi(x)$, $\forall x \in E$.

b) Let $E$, $F$, $G$ be $R$-modules and let $\varphi : E \to F$, $\psi : F \to G$ be $R$-module homomorphisms. Then their composition $\psi \circ \varphi : E \to G$ is also an $R$-module homomorphism. As a consequence, $(\text{End}_R(E), +, \circ)$ is a ring, the unity element being the identity homomorphism $\text{id}_E$.

c) Let $E$ be a submodule of the $R$-module $F$ and let $\psi : F \to G$ be an $R$-module homomorphism. Then the restriction of $\psi$ to $E$, $\psi|_E : E \to G$, is an $R$-module homomorphism. □

A homomorphism is perfectly determined by its values on a generating set:

1.27 Proposition. Let $E$, $F$ be $R$-modules, $S$ a system of generators of $E$ and $\varphi, \psi : E \to F$ module homomorphism. Then $\varphi = \psi$ iff $\varphi|_S = \psi|_S$.

Proof. Suppose $\varphi|_S = \psi|_S$. If $x \in E$, there exist $x_1, \ldots, x_n \in S$ and $r_1, \ldots, r_n \in R$ such that $x = r_1x_1 + \ldots + r_nx_n$. So,

$$\varphi(x) = \varphi(r_1x_1 + \ldots + r_nx_n) = r_1\varphi(x_1) + \ldots + r_n\varphi(x_n) = r_1\psi(x_1) + \ldots + r_n\psi(x_n) = \psi(x).$$ □
1.28 Definition. An $R$-module homomorphism $\varphi : E \to F$ is called an $R$-module isomorphism if there exists a homomorphism $\psi : F \to E$ such that $\varphi \circ \psi = \text{id}_F$ and $\psi \circ \varphi = \text{id}_E$. The $R$-modules $E$ and $F$ are called isomorphic if there exists an $R$-module isomorphism $\varphi : E \to F$. We write in this case $E \cong_R F$ (or $E \cong F$ if it is clear that it is an $R$-module isomorphism). An isomorphism $\varphi : E \to E$ is called an automorphism of $E$.

1.29 Proposition. An $R$-module homomorphism is an isomorphism iff it is bijective.

Proof. Let $\varphi : E \to F$ be a homomorphism. If $\varphi$ is an isomorphism, then it is an invertible map, so it is a bijection. Suppose $\varphi$ is a bijective homomorphism. Then the inverse of the map $\varphi$ exists, $\psi : F \to E$. We must prove that $\psi$ is a homomorphism. Recall that, $\forall y \in F$, $\psi(y) = x$, where $x$ is the unique element in $E$ with $\varphi(x) = y$. Let $y_1, y_2 \in F$ and $x_1, x_2 \in E$ such that $\psi(y_1) = x_1$ and $\psi(y_2) = x_2$. We have $\psi(y_1 + y_2) = x_1 + x_2$, since $\varphi(x_1 + x_2) = \varphi(x_1) + \varphi(x_2) = y_1 + y_2$. So, $\psi(y_1 + y_2) = x_1 + x_2 = \psi(y_1) + \psi(y_2)$. Similarly, $\forall r \in R$, $\forall y \in F$, $\psi(ry) = r \psi(y)$. \hfill $\square$

1.30 Definition. Let $\varphi : M \to N$ be an $R$-module homomorphism. Consider the subset of $M$ defined by:

\[ \ker \varphi := \{ x \in M \mid \varphi(x) = 0 \} = \varphi^{-1}(0) \]

$\ker \varphi$ is called the kernel of $\varphi$. Let $\text{im} \varphi$ denote the image of $\varphi$.

\[ \text{im} \varphi := \{ y \in N \mid \exists x \in M \text{ such that } y = \varphi(x) \} = \{ \varphi(x) \mid x \in M \} \]

1.31 Proposition. If $\varphi : M \to N$ is an $R$-module homomorphism and $M' \leq M$, $N' \leq N$, then $\varphi^{-1}(N')$ is a submodule of $M$, and $\varphi(M)$ is a submodule of $N$. In particular, $\ker \varphi$ is a submodule of $M$, and $\text{im} \varphi$ is a submodule of $N$.

Proof. Let $x, y \in \varphi^{-1}(N')$. Then $\varphi(x - y) = \varphi(x) - \varphi(y) \in N'$, so $x - y \in \varphi^{-1}(N')$. If $r \in R$, then $\varphi(rx) = r \varphi(x) \in N'$, so $rx \in \varphi^{-1}(N')$. 

Let \( z, t \in \varphi(M') \) and \( r \in R \). Then there exist \( x, y \in M' \) such that 
\[ z = \varphi(x), \quad t = \varphi(y). \]
We have \( z - t = \varphi(x) - \varphi(y) = \varphi(x - y) \in \varphi(M') \), and 
\[ rz = r \varphi(x) = \varphi(rx) \in \varphi(M'). \]

1.32 Proposition. Let \( \varphi : M \to N \) be an \( R \)-module homomorphism. 
Then:
\[ a) \ \varphi \text{ is injective iff } \ker \varphi = 0. \]
\[ b) \ \varphi \text{ is surjective iff } \text{Im} \varphi = N. \]

Exercises

1. Prove that the commutativity of the addition of a module is a consequence of the other axioms of the definition of the module.

2. Let \( K \) be a field. Which of the following subsets of the \( K \)-module \( K[X] \) is a \( K \)-submodule?
   \[ a) \ \text{The set of all polynomials of degree 7.} \]
   \[ b) \ \text{The set of all polynomials of degree at most 7.} \]
   \[ c) \ \text{The set of all monic polynomials.} \]
   \[ d) \ \text{The set of all polynomials of degree 1 that have the root 1.} \]
   \[ e) \ \text{The set of all polynomials of even degree.} \]
Which of the sets above is a \( K[X] \)-submodule?

3. Let \( E \) and \( F \) be submodules of the \( R \)-module \( M \). Show that \( E \cup F \) is a submodule of \( M \) iff \( E \subseteq F \) or \( F \subseteq E \).

4. Let \( M \) be an \( R \)-module. Study the properties of the operations \( + \) and \( \cap \) on \( \mathcal{L}_R(M) \) (such as commutativity, associativity, distributivity, existence of neutral elements...).

5. Let \( M \) be an \( R \)-module and let \( A, B, C \subseteq M \). If \( B \subseteq A \), then 
\[ A \cap (B + C) = B + A \cap C \]
(one says that \( \mathcal{L}_R(M) \) is a modular lattice).
6. Determine all submodules of the \( \mathbb{R} \)-module \( \mathbb{R}^2 \).

7. Give an example of a module \( M \) and \( A, B, C \leq M \) such that \( A \cap (B + C) \neq A \cap B + A \cap C \) (i.e., \( \mathcal{L}_R(M) \) is not distributive). (Hint. Try a vector space.)

8. Let \( (G, +) \) be an Abelian group and \( n \in \mathbb{N}^* \) such that \( na = 0 \), \( \forall a \in G \). Then \( G \) has a canonical structure of \( \mathbb{Z}_n \)-module. Is the converse true? Generalization.

9. Identify the Euclidean plane with \( \mathbb{R}^2 \), seen as an \( \mathbb{R} \)-vector space. Which of the following transformations of the plane is a linear transformation (an \( \mathbb{R} \)-module homomorphism) from \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \)?
   - a) The rotation of angle \( \alpha \) around \((0, 0)\).
   - b) The rotation of angle \( \alpha \) around \((0, 1)\).
   - c) The translation by the vector \( v = (x, y) \).
   - d) The symmetry with respect to a line.
   - e) The projection on a line.

10. Let \( R \) be a ring, let \( S \) be a set and \( R^S := \{ \varphi : S \rightarrow R \} \). Define a left \( R \)-module structure on \( R^S \). More generally, let \( M \) be a left \( R \)-module and let \( S \) be a set. Define a left \( R \)-module structure on \( M^S = \{ \varphi : S \rightarrow M \} \).

11. Let \( R \) be a ring. Show that, for any \( R M, R M' \), \( \text{Hom}_R(M, M') \subseteq \text{Hom}_\mathbb{Z}(M, M') \). Give examples of \( R, M, M' \) such that the inclusion is strict.

12. Let \( u : M \rightarrow M' \) be an \( R \)-module homomorphism, \( A, B \leq R M \) and \( A', B' \leq R M' \). Study the validity of the statements:
   - a) \( u(A + B) = u(A) + u(B) \).
   - b) \( u(A \cap B) = u(A) \cap u(B) \).
   - c) \( u^{-1}(A' + B') = u^{-1}(A') + u^{-1}(B') \).
   - d) \( u^{-1}(A' \cap B') = u^{-1}(A') \cap u^{-1}(B') \).

13. Let \( R \) be a commutative ring. Show that \( \text{End}_R(R) \cong R \). More generally, for any \( R M, \text{Hom}_R(R, M) \cong M \).
14. Let \( R \) be a commutative ring. Show that \( R[X] \) is not a finitely generated \( R \)-module.

15. Let \( K \) be a field, \( n \in \mathbb{N}^* \) and \( v_1, \ldots, v_n \in K^n \), where \( v_i = (v_{i1}, \ldots, v_{in}) \). Find necessary and sufficient conditions for \( \{v_1, \ldots, v_n\} \) to be a system of generators for the \( K \)-module \( K^n \).

16. Write 4 distinct homomorphisms from the \( \mathbb{R} \)-module \( \mathbb{R}^2 \) to \( \mathbb{R} \). Which is the general form of such a homomorphism? Generalization.

17. Let \( A \) be an \( R \)-module. Prove that \( A = 0 \iff \text{Hom}_R(A, B) = 0, \forall R B \iff \text{Hom}_R(B, A) = 0, \forall R B \).

18. Does the \( \mathbb{Z} \)-module \( \mathbb{Q} \) have minimal submodules?

19. Let \( V \) be a finite dimensional \( K \)-vector space and let \( u \in \text{End}_K(V) \). Then: \( u \) is injective \( \iff \) \( u \) is surjective \( \iff \) \( u \) is an isomorphism.

20. Give examples of an \( R \)-module \( M \) and of an endomorphism \( \varphi : M \to M \) such that:
   a) \( \varphi \) is injective, but not surjective.
   b) \( \varphi \) is surjective, but not injective.

II.2 Factor modules and the isomorphism theorems

The method used to construct the factor ring (given a ring and an ideal in the ring) can be applied to modules, with minor modifications.

Let \( M \) be a left \( R \)-module and let \( L \) be a submodule in \( M \). Considering only the Abelian group structure on \( M \), \( L \) is a subgroup in \( M \), so we can construct the factor group \( M/L \), which is also an Abelian group. The Abelian group \( M/L \) can be endowed with a natural \( R \)-module structure, inherited from the \( R \)-module structure on \( M \).
II.2 Factor modules and the isomorphism theorems

We briefly recall the construction of the factor group $M/L$. Define on $M$ the equivalence relation (also called “modulo $L$ congruence”) by:

\[ \forall x, y \in M : x \sim y \pmod{L} \iff x - y \in L. \]

An easy check shows that this is indeed an equivalence relation and that the equivalence class of $x \in M$, i.e. the set \{ $y \in M$ | $x \sim y \pmod{L}$ \} is $x + L$, where

\[ x + L := \{ x + l | l \in L \} \] (called “the class of $x$ modulo $L$”)

Define now the set $M/L$ as the set of all equivalence classes:

\[ M/L = \{ x + L | x \in M \} \]

Make $M/L$ into an Abelian group by putting:

\[ (x + L) + (y + L) := (x + y) + L, \forall x, y \in M. \]

One proves that: the operation above is well defined (it does not depend on the representatives of the classes modulo $L$) and $(M/L, +)$ is an Abelian group; the neutral element is $0 + L$ (equal to $l + L$, $\forall l \in L$); $-(x + L) = (-x) + L$, $\forall x \in M$.

Getting back to the module case, define an external operation $R \times (M/L) \rightarrow M/L$: $\forall r \in R$, $\forall x + L \in M/L$, set

\[ r(x + L) := rx + L. \]

This definition is correct: if $r \in R$ and $x, y \in M$, with $x + L = y + L$, then $x - y \in L$, so $r(x - y) \in L$ (since $L \leq_R M$), i.e. $rx - ry \in L$. Thus, $rx + L = ry + L$.

A routine exercise shows that $M/L$ becomes a left $R$-module. For instance, $\forall r, s \in R$, $\forall x \in M$:

\[ (r + s)(x + L) = (r + s)x + L = (rx + sx) + L = (rx + L) + (sx + L) = r(x + L) + s(x + L). \]

2.1 Definition. The $R$-module $M/L$ defined above is called the factor module of $M$ with respect to $L$. The map $\pi : M \rightarrow M/L$, $\pi(x) = x + L$, $\forall x \in M$, is a surjective module homomorphism, called the canonical homomorphism or the canonical surjection.
2.2 Examples. If \( L \leq R M \) and \( \pi : M \to M/L \) is the canonical homomorphism, then
\[
\text{Ker}\pi = \{ x \in M \mid x + L = 0 + L \} = L \quad \text{and} \quad \text{Im}\pi = M/L.
\]
So, any submodule is the kernel of some homomorphism.

2.3 Proposition. Let \( M \) be an \( R \)-module and let \( L \) be a submodule. Then there is a natural one-to-one increasing map between the lattice of the submodules of \( M \) that include \( L \) and the lattice of submodules of \( M/L \), given by
\[
\varphi : \{ A \leq R M \mid L \subseteq A \} \to \mathcal{L}_R(M/L), \quad \varphi(A) := \{ a + L \mid a \in A \},
\]
\( \forall A \leq R M, L \subseteq A \).

Proof. We remark that \( \varphi(A) = \pi(A) \), the image, through the canonical homomorphism \( \pi : M \to M/L \), of the submodule \( A \). So, \( \varphi(A) \leq R M/L \). If \( B \leq R M/L \), then \( \pi^{-1}(B) = \{ x \in M \mid x + L \in B \} \) is a submodule of \( M \) that includes \( L \), and \( \varphi(\pi^{-1}(B)) = B \). So, \( \varphi \) is surjective. The rest of the proof is left to the reader. \( \square \)

2.4 Example. Let \( n \in \mathbb{N} \). Let us find the submodules of the \( \mathbb{Z} \)-module \( \mathbb{Z}/n\mathbb{Z} \) (also known as \( \mathbb{Z}_n \)). The above says that \( \mathcal{L}(\mathbb{Z}/n\mathbb{Z}) \) is in one-to-one correspondence to the submodules of \( \mathbb{Z} \) that include \( n\mathbb{Z} \), which are \( \{ m\mathbb{Z} \mid m \in \mathbb{N}, \ m|n \} \). To \( m\mathbb{Z} \geq n\mathbb{Z} \) corresponds \( m\mathbb{Z}/n\mathbb{Z} \leq \mathbb{Z}/n\mathbb{Z} \). So, \( \mathcal{L}(\mathbb{Z}/n\mathbb{Z}) = \{ m\mathbb{Z}/n\mathbb{Z} \mid m \in \mathbb{N}, \ m|n \} \). Also, \( m\mathbb{Z}/n\mathbb{Z} = \langle m + n\mathbb{Z} \rangle \) is the unique submodule having \( n/m \) elements in \( \mathbb{Z}/n\mathbb{Z}: \{ 0 + n\mathbb{Z}, m + n\mathbb{Z}, \ldots, (n/m - 1)m + n\mathbb{Z} \} \).

The following notions are general concepts in category theory, applied to the category \( R\text{-Mod} \) of the left \( R \)-modules:

2.5 Definition. Let \( \varphi : M \to N \) be an \( R \)-module homomorphism.

a) \( \varphi \) is called a monomorphism if, for any \( _RA \) and any homomorphisms \( u, v : A \to M \), \( \varphi \circ u = \varphi \circ v \) implies \( u = v \).

b) \( \varphi \) is called an epimorphism if, for any \( _RA \) and any homomorphisms \( u, v : N \to A \), \( u \circ \varphi = v \circ \varphi \) implies \( u = v \).
2.6 Proposition. Let \( \varphi : M \rightarrow N \) be a homomorphism. Then:

a) \( \varphi \) is a monomorphism if and only if \( \varphi \) is injective.

b) \( \varphi \) is an epimorphism if and only if \( \varphi \) is surjective.

Proof. a) Suppose \( \varphi \) injective. Let \( R \mathcal{A} \) and \( u, v \in \text{Hom}_R(\mathcal{A}, M) \) such that \( \varphi \circ u = \varphi \circ v \). So, \( \forall a \in \mathcal{A}, \varphi(u(a)) = \varphi(v(a)) \). Since \( \varphi \) is injective, we deduce \( u(a) = v(a) \).

b) Let \( \varphi \) be an epimorphism. Consider the factor module \( N/\text{Im}\varphi \) and \( \pi : N \rightarrow N/\text{Im}\varphi \) (canonical surjection) and \( 0 : N \rightarrow N/\text{Im}\varphi \). We have \( \pi \circ \varphi(x) = \varphi(x) + \text{Im}\varphi = 0 + \text{Im}\varphi = \varphi(0) \), \( \forall x \in M \). So, \( \pi \circ \varphi = \varphi \), i.e. \( \pi = 0 \). This means \( N/\text{Im}\varphi = 0 \iff N = \text{Im}\varphi \). The converse is left to the reader.

2.7 Theorem. (The fundamental isomorphism theorem) Let \( \varphi : M \rightarrow N \) be an \( R \)-module homomorphism. Then \( \frac{M}{\text{Ker}\varphi} \cong \text{Im}\varphi \).

More precisely, there exists a unique isomorphism \( \psi : M/\text{Ker}\varphi \rightarrow \text{Im}\varphi \) such that the following diagram is commutative:

\[
\begin{array}{ccc}
M & \xrightarrow{\varphi} & N \\
\downarrow{\pi} & & \uparrow{i} \\
M/\text{Ker}\varphi & \xrightarrow{\psi} & \text{Im}\varphi
\end{array}
\]

i.e. \( \varphi = i \circ \psi \circ \pi \), where \( i \) is the inclusion and \( \pi \) is the canonical surjection. \( \psi \) is given by \( \psi(x + \text{Ker}\varphi) = \varphi(x), \forall x \in M \).

Proof. \( \psi : M/\text{Ker}\varphi \rightarrow \text{Im}\varphi \) is well defined: if \( x, y \in M \) are such that \( x + \text{Ker}\varphi = y + \text{Ker}\varphi \), then \( x - y \in \text{Ker}\varphi \iff \varphi(x - y) = 0 \iff \varphi(x) = \varphi(y) \). Thus, \( \psi(x + \text{Ker}\varphi) \) does not depend on \( x \), but only on the equivalence class \( x + \text{Ker}\varphi \). Checking that \( \psi \) is a homomorphism is simple.

\( \psi \) is surjective: \( \text{Im}\psi = \{ \psi(x + \text{Ker}\varphi) \mid x \in M \} = \{ \varphi(x) \mid x \in M \} = \text{Im}\varphi \). For the injectivity, we show that \( \text{Ker}\psi = \{ 0 + \text{Ker}\varphi \} \): if \( x \in M \)
with $\psi(x + \ker \varphi) = 0$, then $\varphi(x) = 0$, so $x \in \ker \varphi$, $\iff x + \ker \varphi = 0 + \ker \varphi$.

Also, $t \circ \psi \circ \pi(x) = \psi(x + \ker \varphi) = \varphi(x)$, $\forall x \in M$.

If $\eta : M/\ker \varphi \to \text{Im} \varphi$ is a homomorphism such that $\varphi = t \circ \eta \circ \pi$, then, $\forall x \in M$, $\eta(x + \ker \varphi) = t\eta \pi(x) = \varphi(x)$, so $\eta = \psi$.

2.8 Remark. A typical use for the isomorphism theorem is as follows: suppose $B \leq R A$ and we want to prove that $A/B \cong C$. One looks for a surjective homomorphism $\varphi : A \to C$, with $\ker \varphi = B$. The isomorphism theorem provides then the required isomorphism. The following corollaries illustrate this technique (which is used also for groups, rings, algebras, ...).

2.9 Corollary. Let $R M$ and $E, F \leq M$ such that $E \subseteq F$. Then $F/E$ is a submodule of $M/E$ and:

$$\frac{M/E}{F/E} \cong \frac{M}{F} \quad \text{(R-module isomorphism).}$$

Proof. Since $F/E = \{x + E \mid x \in F\}$, $F/E \subseteq M/E = \{x + E \mid x \in M\}$. Let $\varphi : M/E \to M/F$, $\varphi(x + E) = x + F$, $\forall x \in M$. The map $\varphi$ is well-defined: if $x, y \in M$, with $x + E = y + E$, then $x - y \in E$. So, $x - y \in F$ and $x + F = y + F$. It easy to see that $\varphi$ is a surjective module homomorphism. $\ker \varphi = \{x + E \mid x \in M, x + F = 0 + F\} = F/E$. Apply now the fundamental isomorphism theorem.

2.10 Corollary. Let $R M$ and $E, F \leq R M$. Then

$$\frac{E + F}{F} \cong \frac{E}{E \cap F}.$$  

Proof. Let $\varphi : F \to (E + F)/E$, $\varphi(x) = x + E$, $\forall x \in F$. The mapping $\varphi$ is a module homomorphism, being the restriction of the canonical homomorphism $E + F \to (E + F)/E$ to the submodule $F$ of $E + F$. Moreover, $\varphi$ is surjective: $\forall (e + f) + E \in (E + F)/E$, with $e \in E, f \in F$, $(e + f) + E = f + E = \varphi(f)$.
II.2 Factor modules and the isomorphism theorems

\[ \text{Ker}\varphi = \{ x \in F \mid x + E = 0 + E \} = \{ x \in F \mid x \in E \} = E \cap F. \]

Apply the fundamental isomorphism theorem, we get \( F/(E \cap F) \cong (E + F)/E. \)

The following result is often used in module theory arguments:

2.11 Proposition. Let \( \varphi : E \to F \) and \( \psi : E \to G \) be \( R \)-module homomorphisms, with \( \varphi \) surjective. If \( \text{Ker}\varphi \subseteq \text{Ker}\psi \), then:

a) “\( \psi \) factorizes through \( \varphi \)”: there exists a unique homomorphism \( \eta : F \to G \) such that \( \psi = \eta \circ \varphi \).

b) \( \eta \) is injective if and only if \( \text{Ker}\varphi = \text{Ker}\psi \).

Proof. a) Let \( y \in F \). Since \( \varphi \) is surjective, there exists \( x \in E \) with \( \varphi(x) = y \). Define \( \eta(y) := \psi(x) \). The definition is independent on the choice of \( x \in E \) with \( \varphi(x) = y \). Indeed, if \( x, x' \in E \) such that \( \varphi(x) = \varphi(x') = y \), then \( x - x' \in \text{Ker}\varphi \subseteq \text{Ker}\psi \), so \( \psi(x) = \psi(x') \). Also, \( \eta \) is a homomorphism (standard check) and, \( \forall x \in E, \ \eta(\varphi(x)) = \psi(x) \). If \( \eta' \in \text{Hom}_R(F, G) \) has the property that \( \eta' \circ \varphi = \psi \), then \( \eta(\varphi(x)) = \eta'(\varphi(x)) \), \( \forall x \in E \). Since \( \varphi \) is surjective, \( \eta = \eta' \).

b) Suppose \( \eta \) is injective. Let \( x \in E \) with \( \psi(x) = 0 \). Then \( \eta(\varphi(x)) = 0 \), so \( \varphi(x) = 0 \), i.e. \( x \in \text{Ker}\varphi \). Therefore \( \text{Ker}\varphi = \text{Ker}\psi \). Suppose now that \( \text{Ker}\varphi = \text{Ker}\psi \) and take \( y \in F \) with \( \eta(y) = 0 \). Then \( \exists x \in E \) with \( \varphi(x) = y \); so, \( \eta(\varphi(x)) = \psi(x) = 0 \), which means \( x \in \text{Ker}\psi \). Since \( \text{Ker}\varphi = \text{Ker}\psi \), \( x \in \text{Ker}\varphi \), so \( \varphi(x) = y = 0 \). \( \square \)
Exercises

1. Prove that every $R$-module homomorphism $u : M \to N$ can be written as $u = v \circ w$, where $v$ is an injective homomorphism and $w$ is a surjective homomorphism.

2. Let $M$ be a finitely generated $R$-module. Show that there exists $n \in \mathbb{N}$ and a surjective homomorphism $\varphi : R^n \to M$. Deduce that, if $M$ is cyclic (generated by one element), then $M$ is isomorphic to a $R$-module of the form $R/I$, where $I \leq R$.

3. Prove that every factor module of a finitely generated module is finitely generated.

4. Let $m, n \in \mathbb{N}^*$. Prove that:
   a) $\text{Hom}_Z(\mathbb{Z}_m, \mathbb{Z}) = 0$ and $\text{Hom}_Z(\mathbb{Z}, \mathbb{Z}_n) \cong \mathbb{Z}_n$.
   b) $\text{Hom}_Z(\mathbb{Z}_m, \mathbb{Z}_n) \cong \mathbb{Z}_d$, where $d = \text{GCD}(m, n)$.

5. Let $m, n \in \mathbb{N}$. Under what conditions the Abelian group $(\mathbb{Z}_m, +)$ has a $\mathbb{Z}_n$-module structure? Does there exist a $\mathbb{Z}_n$-module structure on $(\mathbb{Z}, +)$?

6. Determine all the submodules in: $\mathbb{Z}_6, \mathbb{Z}_8, \mathbb{Z}_{24}$. Generalize the result.

7. Prove the following version of the fundamental isomorphism theorem: Let $\varphi : M \to N$ be an $R$-module homomorphism and let $L \leq R M$ such that $\text{Ker} \varphi \subseteq L$. Then there exists a canonical isomorphism $\frac{M}{L} \cong \frac{\varphi(M)}{\varphi(L)}$.

8. An $R$-module is called simple if it has no submodules other than zero or itself. Show that $M$ is simple if and only if it is isomorphic to a module of the type $R/I$, where $I$ is a maximal left ideal in $R$. Write all (up to isomorphism) simple $\mathbb{Z}$-modules and all simple $K[X]$-modules, where $K$ is a field.
II.3 Direct sums and products. Exact sequences

9. Is it true that any intersection of finitely generated submodules is a finitely generated submodule? (Hint. Let $K$ be a field, $S = K[X_n; n \in \mathbb{N}]$ and $T = S/I$, where $I$ is the ideal of $S$ generated by \{(X_1 - X_2)X_i \mid i \geq 3\}. Let $\xi_1, \xi_2$ be the images of $X_1, X_2$ in $T$. Then $T\xi_1 \cap T\xi_2$ is not a finitely generated $T$-module.)

II.3 Direct sums and products. Exact sequences

The methods of constructing new structures (from a given collection of structures) are extremely important. In the case of modules (and also groups, rings...) some of the most used constructions are the direct product and the direct sum.

We begin with the case of two $R$-modules $M_1$ and $M_2$. The Cartesian product $M_1 \times M_2 = \{(x_1, x_2) \mid x_1 \in M_1, x_2 \in M_2\}$ is made into an Abelian group by defining:

$$(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2), \forall (x_1, x_2), (y_1, y_2) \in M_1 \times M_2.$$

$M_1 \times M_2$ is in fact the direct product of the groups $(M_1, +)$ and $(M_2, +)$. The Abelian group $M_1 \times M_2$ becomes a left $R$-module by defining:

$$r(x_1, x_2) := (rx_1, rx_2), \forall r \in R, \forall (x_1, x_2) \in M_1 \times M_2.$$

It is trivial to check the $R$-module axioms.

Define the homomorphisms $\pi_1 : M_1 \times M_2 \to M_1$ and $\pi_2 : M_1 \times M_2 \to M_2$, by:

$$\pi_1(x_1, x_2) := x_1 \text{ and } \pi_2(x_1, x_2) := x_2, \forall (x_1, x_2) \in M_1 \times M_2.$$

The $R$-module $M_1 \times M_2$, together with the homomorphisms $\pi_1$ and $\pi_2$, is called the direct product of $M_1$ and $M_2$. The homomorphisms $\pi_1$ and $\pi_2$ are called the canonical projections of the direct product.
$M_1 \times M_2$. The direct product satisfies the following universality property:

**3.1 Theorem.** (the universality property of the direct product) Let $M_1$ and $M_2$ be $R$-modules. Then, for any $R$-module $E$ and any homomorphisms $v_1 : E \to M_1$, $v_2 : E \to M_2$, there exists a unique homomorphism $v : E \to M_1 \times M_2$ such that $v_1 = \pi_1 \circ v$ and $v_2 = \pi_2 \circ v$ (the diagram below is commutative, for $i \in \{1,2\}$):

\[
\begin{array}{c}
M_1 \times M_2 \\
\downarrow \quad \downarrow \pi_i \\
M_i \\
\uparrow \quad \uparrow v_i \\
E
\end{array}
\]

**Proof.** We prove first the uniqueness. Let $v : E \to M_1 \times M_2$ be a homomorphism with the required property. Let $e \in E$ and suppose $v(e) = (x_1, x_2) \in M_1 \times M_2$. We have $x_1 = \pi_1(v(e)) = (\pi_1 \circ v)(e) = v_1(e)$ and, likewise, $x_2 = v_2(e)$. So, if there exists $v$ as required, then $v(e) = (v_1(e), v_2(e))$, $\forall e \in E$. On the other hand, the map $v$ defined in this way is clearly a module homomorphism.

The construction above can be generalized for an arbitrary family (possibly infinite) of modules. Let $I$ be a set (seen as a set of indices) and let $(M_i)_{i \in I}$ be a family of $R$-modules indexed by $I$. Recall that the Cartesian product of this family is

\[
\prod_{i \in I} M_i = \{ f : I \to \bigcup_{i \in I} M_i \mid f(i) \in M_i, \forall i \in I \}.
\]

An element $f \in \prod_{i \in I} M_i$ is usually written as $(x_i)_{i \in I}$ or $(x_i)_I$ (where we denoted $f(i) = x_i \in M_i, \forall i \in I$). Define an operation of *addition* on $\prod_{i \in I} M_i$ by:

\[
\forall f, g \in \prod_{i \in I} M_i, (f + g)(i) := f(i) + g(i), \forall i \in I;
\]

\[9\] So, $v_1$ factorizes through $\pi_1$ and $v_2$ factorizes through $\pi_2$.\]
With the alternate notation: \( \forall (x_i)_I, (y_i)_I \in \prod_{i \in I} M_i, \)
\( (x_i)_I + (y_i)_I := (x_i + y_i)_I. \)

\( \prod_{i \in I} M_i \) is an Abelian group with respect to this operation.

Define the external operation: \( \forall r \in R, \forall f \in \prod_{i \in I} M_i, \)
\( (rf)(i) := rf(i), \forall i \in I; \)

With the alternate notation: \( \forall r \in R, \forall (x_i)_I \in \prod_{i \in I} M_i, \)
\( r(x_i)_I := (rx_i)_I. \)

Proving that \( \prod_{i \in I} M_i \) becomes an \( R \)-module with the operations above is routine: for instance, \( \forall r \in R, \forall f, g \in \prod_{i \in I} M_i, \forall i \in I, \) we have
\( (r(f + g))(i) = r(f + g)(i) = r(f(i) + g(i)) = rf(i) + rg(i) = (rf + rg)(i), \)
which shows that \( r(f + g) = rf + rg. \)

For any \( j \in I, \) define the homomorphisms \( \pi_j : \prod_{i \in I} M_i \to M_j \) by:
\( \pi_j(f) := f(j), \forall f \in \prod_{i \in I} M_i; \)

using the notation \( (x_i)_I : \pi_j((x_i)_I) := x_j, \forall (x_i)_I \in \prod_{i \in I} M_i. \)

The \( R \)-module \( \prod_{i \in I} M_i, \) together with the family of homomorphisms \( (\pi_i)_{i \in I}, \) is called the direct product of the family of modules \( (M_i)_{i \in I}. \) The homomorphisms \( \pi_i, i \in I, \) are called the canonical projections. The direct product satisfies a universality property:

**3.2 Theorem.** (the universality property of the direct product) Let \( (M_i)_{i \in I} \) be family of \( R \)-modules. Then, for any \( R \)-module \( E \) and any family of homomorphisms indexed by \( I, v_i : E \to M_i, i \in I, \) there exists a unique homomorphism \( \nu : E \to \prod_{i \in I} M_i \) such that \( v_i = \pi_i \circ \nu, \forall i \in I. \)
II. Modules

Proof. It is basically the same as for two modules. Suppose there exists a homomorphism \( \nu : E \to \prod_{i \in I} M_i \) with the desired property. Let \( e \in E \) and denote \( \nu(e) \) by \( f \in \prod_{i \in I} M_i \). For any \( i \in I \), \( f(i) = \pi_i(f) = \pi_i(\nu(e)) = (\pi_i \circ \nu)(e) = \nu_i(e) \). So, if such a homomorphism \( \nu \) exists, then \( \nu(e)(i) = \nu_i(e) \), \( \forall e \in E \). On the other hand, \( \nu \) defined this way is a module homomorphism. \( \square \)

The direct product of the family \((M_i)_{i \in I}\) is denoted also \( \times_{i \in I} M_i \). For a finite family \( M_1, \ldots, M_n \), the notation is \( M_1 \times \ldots \times M_n \) or \( \prod_{i=1}^n M_i \). If the modules of the family \((M_i)_{i \in I}\) are all equal to some module \( M \), the direct product of the family is denoted by \( M^I \) (or \( M^n \) if \( I \) is finite, having \( n \) elements).

A remarkable fact is that the universality property of the direct product characterizes the direct product up to a unique isomorphism:

3.3 Theorem. Let \((M_i)_{i \in I}\) be a family of \( R \)-modules. Suppose that an \( R \)-module \( P \), and the family of homomorphisms indexed by \( I \), \( p_i : P \to M_i \), \( i \in I \), satisfies the property:

“For any \( R \)-module \( E \) and any family of homomorphisms indexed by \( I \), \( \nu_i : E \to M_i \), \( i \in I \), there exists a unique homomorphism \( \nu : E \to P \) such that \( \nu_i = p_i \circ \nu \), \( \forall i \in I \).” \( \text{(U)} \)

If the \( R \)-module \( Q \), together with the family of homomorphisms \( q_i : Q \to M_i \), \( i \in I \), satisfies also property \( \text{(U)} \) above, then there exists a unique isomorphism \( \varphi : P \to Q \) such that \( p_i = q_i \circ \varphi \), \( \forall i \in I \).

Proof. In \( \text{(U)} \), set \( E = Q \) and \( \nu_i = q_i \), \( \forall i \in I \). We obtain a (unique) homomorphism \( \psi : Q \to P \) such that \( q_i = p_i \circ \psi \). Since \( Q \) satisfies also

\[10\] This is exactly the universality property of the direct product.
II.3 Direct sums and products. Exact sequences

(U), apply (U) for \( P \) with the homomorphisms \( (p_i)_{i \in I} \), there exists a unique homomorphism \( \varphi : P \to Q \), such that \( p_i = q_i \circ \varphi \). We show that \( \varphi \) and \( \psi \) are inverse one to each other. Indeed, \( \varphi \circ \psi : Q \to Q \) satisfies \( q_i \circ (\varphi \circ \psi) = (q_i \circ \varphi) \circ \psi = p_i \circ \psi = q_i, \forall i \in I \); but the homomorphism \( \text{id}_Q : Q \to Q \) has the same property: \( q_i \circ \text{id}_Q = q_i, \forall i \in I \). By uniqueness, guaranteed by (U), we get \( \varphi \circ \psi = \text{id}_Q \). Similarly, \( \psi \circ \varphi = \text{id}_P \). \( \square \)

3.4 Remark. The universality property of the direct product is articulated only in terms of objects (modules, in this case) and homomorphisms, without involving any elements of the underlying sets of the modules. For this reason, the universality property above is taken as a definition of the direct product in a category \( \mathcal{C} \) (replace “module” with “object in \( \mathcal{C} \)” and “module homomorphism” with “morphism in \( \mathcal{C} \)”). The direct product of a family of modules in \( \mathcal{C} \) (if it exists) is uniquely determined up to an isomorphism: the proof of the proposition above is valid in any category. Theorem 3.2 says that in \( R\text{-Mod} \) the direct product of an arbitrary family of objects \( (M_i)_{i \in I} \) exists, namely \( (\prod_{i \in I} M_i, (\pi_i)_{i \in I}) \).

Dualizing (“reversing the arrows”) in the universality property of the direct product, one obtains the universality property of the direct sum (taken as a definition of this notion):

3.5 Definition. Let \( (M_i)_{i \in I} \) be a family of \( R \)-modules. An \( R \)-module \( S \), together with a family of homomorphisms \( (\sigma_i)_{i \in I}, \sigma_i : M_i \to S \), is called a direct sum of the family \( (M_i)_{i \in I} \) if, for any module \( E \) and any family of homomorphisms \( (\nu_i)_{i \in I}, \nu_i : M_i \to E \), there exists a unique homomorphism \( \nu : S \to E \) such that \( \nu \circ \sigma_i = \nu_i, \forall i \in I \) (the diagram below is commutative, \( \forall i \in I \)):
The homomorphisms \((\sigma_i)_{i \in I}\) are called the **canonical injections** of the direct sum \((S, (\sigma_i)_{i \in I})\).

3.6 **Remark.** The canonical injections are indeed injective. For a fixed \(j \in I\), in the definition above set \(E = M_j\) and \(v_i : M_i \to M_j\) defined as follows: \(v_i = 0\) if \(i \neq j\) and \(v_j = \text{id}_{M_j}\). The homomorphism \(v : S \to M_j\), given by the definition, satisfies \(v \circ \sigma_j = \text{id}\), so \(\sigma_j\) is injective.

The direct sum is also called **coproduct** (a categorical name, emphasizing the duality with the product).

We now construct an object satisfying the definition of the direct sum.

If \(f \in \prod_{i \in I} M_i\), recall that the **support** of \(f\) is the set \(\text{Supp} f := \{i \in I \mid f(i) \neq 0\}\). Consider the following subset of \(\prod_{i \in I} M_i\):

\[
\coprod_{i \in I} M_i := \{f \in \prod_{i \in I} M_i \mid \text{Supp} f \text{ is finite}\}.
\]

We claim that \(\coprod_{i \in I} M_i\) is a submodule of \(\prod_{i \in I} M_i\). Indeed, if \(f, g \in \coprod_{i \in I} M_i\), then \(\text{Supp}(f + g) \subseteq \text{Supp} f \cup \text{Supp} g\), so \(f + g\) has finite support. If \(r \in R\), then \(\text{Supp}(rf) \subseteq \text{Supp} f\). Identify \(f \in \coprod_{i \in I} M_i\) with the “family of elements” \((x_i)_{i \in I}\), where \(x_i = f(i) \in M_i\), \(\forall i \in I\); the family \((x_i)_{i \in I}\) has finite support.

Define, \(\forall j \in I\), \(\sigma_j : M_j \to \coprod_{i \in I} M_i\) by:

\[
(\sigma_j(x))(i) := \begin{cases} 
0, & \text{if } i \neq j \\
x, & \text{if } i = j 
\end{cases}.
\]

In other words, if \(\sigma_j(x) = (x_i)_{i \in I}\), then \(x_j = x\), and \(x_i = 0\), \(\forall i \in I, i \neq j\). It is easy to check that \(\sigma_j\) is a homomorphism, \(\forall j \in I\). Note that \(\forall (x_i)_{i \in I} \in \coprod_{i \in I} M_i\), we have

\[
(x_i)_{i \in I} = \sum_{i \in I} \sigma_i(x_i)
\]

(the sum is finite, \(i\) runs only on \(\text{Supp}((x_i)_{i \in I})\), which is finite).
3.7 Theorem. Let \((M_i)_{i \in I}\) be a family of \(R\)-modules. Then the module \(\prod_{i \in I} M_i\), together with the homomorphisms \((\sigma_i)_{i \in I}\), is a direct sum of the family \((M_i)_{i \in I}\).

Proof. Let \(E\) be an \(R\)-module and let \(v_i : M_i \to E\), \(\forall i \in I\). If \(v : \prod_{i \in I} M_i \to E\) is a homomorphism with \(v \circ \sigma_i = v_i\), \(\forall i \in I\), then, \(\forall (x_i)_{i \in I} \in \prod_{i \in I} M_i\), we have:

\[
v((x_i)_{i \in I}) = v(\sum_{i \in I} \sigma_i(x_i)) = \sum_{i \in I} (v \circ \sigma_i)(x_i) = \sum_{i \in I} v_i(x_i).\]

This shows that the homomorphism \(v\) is uniquely determined by the condition \(v \circ \sigma_i = v_i\), \(\forall i \in I\). If we define \(v\) in this way, \(v\) is indeed a homomorphism:

\[
\begin{align*}
v((x_i)_{i \in I} + (y_i)_{i \in I}) &= \sum_{i \in I} v_i(x_i + y_i) = \sum_{i \in I} v_i(x_i) + v_i(y_i) \\
v(r(x_i)_{i \in I}) &= \sum_{i \in I} v_i(rx_i) = r \sum_{i \in I} v_i(x_i) = rv((x_i)_{i \in I}).
\end{align*}
\]

The module \(\prod_{i \in I} M_i\) is often denoted \(\bigoplus_{i \in I} M_i\). For a finite family \(M_1, \ldots, M_n\), the notation is \(M_1 \oplus \ldots \oplus M_n\) or \(\bigoplus_{i=1}^n M_i\). If the modules in the family \((M_i)_{i \in I}\) are all equal to the same module \(M\), \(\bigoplus_{i \in I} M_i\) is denoted by \(M^{(I)}\). In order to avoid the confusion with the notion of internal direct sum of submodules (see 3.11), \(\prod_{i \in I} M_i\) is sometimes called external direct sum.

As in the case of the direct product, the universality property characterizes the direct sum of a family of modules up to an isomorphism. The reader is invited to formulate this precisely and prove it, as in the case of the direct product.

3.8 Remark. If the set of indices \(I\) is finite, the module direct sum \(\bigoplus_{i \in I} M_i = \prod_{i \in I} M_i\) coincides with the module direct product \(\prod_{i \in I} M_i\) in 3.2. Nevertheless, the direct sum is a couple \((\bigoplus_{i \in I} M_i, (\sigma_i)_{i \in I})\) and is not the same with the direct product \((\prod_{i \in I} M_i, (\pi_i)_{i \in I})\).
3.9 Example. Let $R$ be a commutative ring and let $R[X]$ be the polynomial ring in the indeterminate $X$. As an $R$-module, $R[X]$ is isomorphic with the direct sum $R^{(\mathbb{N})}$ of a countable family of copies of $R$.

We defined the notion of sum of a family $(L_i)_{i \in I}$ of submodules of $R \cdot M$. The direct sum of the family $(L_i)_{i \in I}$, where $L_i$ are seen as modules by themselves can be also constructed. When are these sums “the same”?

3.10 Proposition. Let $M$ be an $R$-module and let $(L_i)_{i \in I}$ be a family of submodules of $M$. Let $L$ denote their sum, $L := \sum_{i \in I} L_i$, and let $\eta_i : L_i \to L$ be the inclusion homomorphisms. The following statements are equivalent:

a) $L$ is the external direct sum of the modules $(L_i)_{i \in I}$ and the $(\eta_i)_{i \in I}$ are the canonical injections.

b) For any $x \in L$, there exists a unique family $(x_i)_{i \in I}$ with $x_i \in L_i$, $\forall i \in I$, having finite support, such that $x = \sum_{i \in I} x_i$.

c) For any $j \in I$, $L_j \cap \left( \sum_{i \in I \setminus \{j\}} L_i \right) = 0$.

d) For any family $(x_i)_{i \in I}$, with $x_i \in L_i$, $\forall i \in I$, having finite support, such that $\sum_{i \in I} x_i = 0$, we have $x_i = 0$, $\forall i \in I$.

Proof. $a) \Rightarrow b)$ Let $\bigcup_{i \in I} L_i$ be the direct sum constructed as in 3.7, with $(\sigma_i)_{i \in I}$ the canonical injections; let $\varphi : \bigcup_{i \in I} L_i \to L$ be the unique isomorphism with $\varphi \circ \sigma_i = \eta_i$, $\forall i \in I$. For any $x \in L$, there exists $(x_i)_{i \in I} \in \bigcup_{i \in I} L_i$ such that $\varphi((x_i)_{i \in I}) = x$. So,

$$x = \varphi((x_i)_{i \in I}) = \varphi(\sum_{i \in I} \sigma_i(x_i)) = \sum_{i \in I} (\varphi \circ \sigma_i)(x_i) = \sum_{i \in I} \eta_i(x_i) = \sum_{i \in I} x_i,$$

which means that $x$ is the sum of the family $(x_i)_{i \in I}$, $x_i \in L_i$, $\forall i \in I$, having finite support. If $(y_i)_{i \in I}$ is another family having finite support, with $y_i \in L_i$, $\forall i \in I$, such that $\sum_{i \in I} x_i = \sum_{i \in I} y_i$, then $\varphi((x_i)_{i \in I}) = \varphi((y_i)_{i \in I})$. Since $\varphi$ is an isomorphism, $(x_i)_{i \in I} = (y_i)_{i \in I}$. 


II.3 Direct sums and products. Exact sequences

b)⇒c) Let \( x_j \in L_j \cap \left( \sum_{i \in I \setminus \{j\}} L_i \right) \). Then there exists a family of finite support, \((x_i)_{i \in I \setminus \{j\}}, x_i \in L_i, \forall i \in I \setminus \{j\}\), such that \( x_j = \sum_{i \in I \setminus \{j\}} x_i \). We obtain that 0 is the sum of the finite support family \((y_i)_{i \in I}\), where \( y_i = x_j, \forall i \neq j \) and \( y_j = -x_j \). Since 0 has a unique writing as a sum of a finite support family (evidently, 0 is the sum of the family \((0)_{i \in I}\)), we obtain \( x_i = 0, \forall i \in I \). So, \( x_j = 0 \).

c)⇒d) Let \((x_i)_{i \in I}, x_i \in L_i, \forall i \in I\), be such that \( \sum_{i \in I} x_i = 0 \). Let \( j \in I \).
Then \( x_j \in L_j \); since \( x_j = \sum_{i \in I \setminus \{j\}} (-x_i) \), we get \( x_j \in \left( \sum_{i \in I \setminus \{j\}} L_i \right) \). So, \( x_j = 0 \).

d)⇒a) We show that \((L, (\eta_i)_{i \in I})\) satisfies the definition 3.5. Let \( R \times E \) and let a family of homomorphisms \((v_i)_{i \in I}, v_i : M_i \to E\). If \( x \in L \), there exists a family of finite support \((x_i)_{i \in I} \) \((x_i \in L_i, \forall i \in I)\), such that \( x = \sum_{i \in I} x_i \). From d), we obtain that this family is unique: if \( \sum_{i \in I} x_i = \sum_{i \in I} y_i \), with \( y_i \in L_i, \forall i \in I \), then \( \sum_{i \in I} (x_i - y_i) = 0 \), so \( x_i = y_i, \forall i \in I \).

Define the homomorphism \( \varphi : L \to E \) by: \( \forall x \in L \), put \( \varphi(x) := \sum_{i \in I} v_i(x_i) \), where \((x_i)_{i \in I}\) is the unique family of finite support with \( x_i \in L_i, \forall i \in I \), and \( x = \sum_{i \in I} x_i \). The map \( \varphi \) thus defined is an \( R \)-module homomorphism: if \( x, y \in L \), and \((x_i)_{i \in I}, (y_i)_{i \in I}\) are the unique families of finite support with \( x_i, y_i \in L_i, \forall i \in I \) and such that \( x = \sum_{i \in I} x_i, \ y = \sum_{i \in I} y_i \), then \( x + y = \sum_{i \in I} x_i + \sum_{i \in I} y_i = \sum_{i \in I} (x_i + y_i) \), where \((x_i + y_i)_{i \in I}\) has finite support and \( x_i + y_i \in L_i, \forall i \in I \). So,
\[
\varphi(x + y) = \sum_{i \in I} v_i(x_i + y_i) = \sum_{i \in I} v_i(x_i) + \sum_{i \in I} v_i(y_i) = \varphi(x) + \varphi(y).
\]
Similarly one sees that \( \varphi(rx) = r \varphi(x), \forall r \in R, \forall x \in L \). If \( j \in I \) and \( x_j \in L_j \), then \( (\varphi \circ \eta_j)(x_j) = \varphi(x_j) = v_j(x_j) \), which shows that \( \varphi \circ \eta_j = v_j, \forall j \in I \). We must show that \( \varphi \) is unique with this property. Let
ψ : L → E be a homomorphism with ψ ◦ η_i = v_i, ∀i ∈ I, and let x ∈ L. Then x = \sum_{i \in I} x_i, x_i ∈ L_i, ∀i ∈ I and:

\[ \psi(x) = \psi(\sum_{i \in I} x_i) = \sum_{i \in I} \psi(x_i) = \sum_{i \in I} \psi(\eta_i(x_i)) = \sum_{i \in I} v_i(x_i) = \varphi(x). \]

3.11 Definition. Let \((L_i)_{i \in I}\) be a family of submodules of an \(R\)-module \(M\). We say that “the sum of the family \((L_i)_{i \in I}\) is a direct (internal) sum” if and only if one of the equivalent statements of the proposition above is satisfied. In this situation, one also says that \((L_i)_{i \in I}\) is an independent family of submodules. A submodule \(A\) of \(M\) is called a direct summand of \(M\) if there exists \(0 \neq B \leq_R M\) such that \(M = A \oplus B\) (a submodule \(B\) having this property is called a complement of \(A\)).

3.12 Corollary. Let \(M\) be an \(R\)-module and let \(A, B \leq_R M\). The following statements are equivalent:

a) \(M = A \oplus B\).

b) \(M = A + B\) and \(A \cap B = 0\).

c) \(∀x \in M, there exists a unique pair (a, b) ∈ A \times B\) such that \(x = a + b\). ⊠

3.13 Examples. a) If \(V\) is a \(K\)-vector space, then any vector subspace \(U\) of \(V\) is a direct summand in \(V\).

b) If \(R\) is a domain, then \(R \cdot R\) has no nonzero submodule that is a direct summand. Indeed, for any nonzero submodules \(A\) and \(B\) of \(R \cdot R\), we have \(A \cap B \neq 0\): for \(0 \neq a \in A\) and \(0 \neq b \in B\), \(0 \neq ab \in A \cap B\).

3.14 Remark. a) The notions of (internal) direct sum of submodules and external direct sum are very close: if \(S\) is a direct sum of the modules \((M_i)_{i \in I}\), with canonical injections \((\sigma_i)_{i \in I}\), then \(S\) is the internal direct sum of its submodules \((\text{Im} \sigma_i)_{i \in I}\), and \(M_i \cong \text{Im} \sigma_i, \forall i \in I\). It is usual to identify \(M_i\) with \(\text{Im} \sigma_i\) and call \(S\) the direct sum of the modules \(M_i, i \in I\).
II.3 Direct sums and products. Exact sequences

b) If $M = \bigoplus_{i \in I} M_i$ is the direct sum of the modules $(M_i)_{i \in I}$, with canonical injections $(\sigma_i)_{i \in I}$, then, $\forall j \in I$, the mapping $p_j : M \to M_j$, $x \mapsto x_j$, (where $x = \sum_{i \in I} \sigma_i(x_i)$, with $x_i \in M_i$, $\forall i \in I$) is a module homomorphism, called the canonical projection on $M_j$. Considering $M$ as a submodule of the direct product $\prod_{i \in I} M_i$, $p_j$ is exactly the restriction to $M$ of the canonical projection $\pi_j$ of the direct product. Furthermore, $p_j \circ \sigma_j$ is the identity of $M_j$:

The direct summands of an $R$-module $M$ are closely related with the idempotents of the ring $\text{End}_R(M)$. (An element $e$ of a ring $S$ is called an idempotent if $e^2 = e$).

3.15 Proposition. Let $M$ be an $R$-module. If $M = A \oplus B$, then $p := i_A \circ p_A$ (where $p_A : M \to A$ is the canonical projection and $i_A : A \to M$ is the canonical injection) is an idempotent of $\text{End}_R(M)$ and $\text{Im} \ p = A$, $\text{Ker} \ p = B$.

Conversely, let $p \in \text{End}_R(M)$ be an idempotent. Then $M = \text{Im} \ p \oplus \text{Ker} \ p$.

Proof. For any $x \in M$, there exist unique $a \in A$, $b \in B$, such that $x = a + b$. Then $p(x) = a$. We have: $p(p(x)) = p(a) = a = p(x)$. So, $p \circ p = p^2 = p$. We have $p(x) = 0$ iff $x = 0 + b$, with $b \in B$, i.e. $\text{Ker} \ p = B$. Evidently, $\text{Im} \ p = A$.

Suppose now that $p \in \text{End}_R(M)$ and $p^2 = p$. For any $x \in M$, write $x = p(x) + (x - p(x))$.

We have $p(x) \in \text{Im} \ p$ and $p(x) - p(x) = p(x) - p^2(x) = 0$, so $x - p(x) \in \text{Ker} \ p$. Thus, $M = \text{Im} \ p + \text{Ker} \ p$. If $x \in \text{Im} \ p \cap \text{Ker} \ p$, then $p(x) = 0$ and $x = p(y)$, with $y \in M$. Therefore, $x = p(y) = p^2(y) = p(x) = 0$ and $\text{Im} \ p \cap \text{Ker} \ p = 0$, which means that $M = \text{Im} \ p \oplus \text{Ker} \ p$. \qed

3.16 Proposition. Let $M$ and $N$ be $R$-modules and let $u : M \to N$ and $v : N \to M$ be homomorphisms such that $v \circ u = \text{id}_M$. Then
\( N = \text{Im} \ u \oplus \ker v. \) In particular, \( M \) is isomorphic to a direct summand of \( N. \)

**Proof.** Let \( p = u \circ v : N \to N. \) Since \( p^2 = u \circ v \circ u \circ v = u \circ \text{id} \circ v = u \circ v = p, \)

\( N = \text{Im} \ p \oplus \ker p \) by 3.15. Because \( v \) is surjective, \( \text{Im} \ p = \text{Im} \ u \circ v = \text{Im} \ u. \) Also, \( \ker u \circ v = \{ x \in N \mid u(v(x)) = 0 \} \)

\( = \{ x \in N \mid v(x) = 0 \} = \ker v, \) since \( u \) is injective. \( \square \)

**3.17 Definition.** Let \( v_i : M_i \to N_i \ (i \in I) \) be a family of \( R \)-module homomorphisms. We define the **direct product** and the **direct sum** of the family of homomorphisms \( (v_i)_{i \in I}. \)

Let \( (\prod M_i, (\pi_i)_{i \in I}) \) (respectively \( (\prod N_i, (\rho_i)_{i \in I}) \) be the direct product of the family \( (M_i)_{i \in I} \) (respectively \( (N_i)_{i \in I} \)). For any \( j \in I, \)

\( \pi_j \circ v_j : \prod M_i \to N_j \) is a homomorphism. The universality property of the direct product (3.2), produces a unique homomorphism

\[ \prod M_i \xrightarrow{\pi_j} M_j \]

\[ \downarrow \quad \downarrow v_j \]

\[ \prod N_i \xrightarrow{\rho_j} N_j \]

\( v : \prod M_i \to \prod N_i \) such that \( \rho_j \circ v = v_j \circ \pi_j, \ \forall j \in I: \)

The homomorphism \( v \) is called the **direct product of the family of homomorphisms** \( (v_i)_{i \in I} \) and is usually denoted by \( \prod_{i \in I} v_i \) or \( \times_{i \in I} v_i. \) If \( I = \{ 1, \ldots, n \}, \) the notations are \( \prod_{i=1}^n v_i \) or \( v_1 \times \ldots \times v_n. \)

For any \( x = (x_i)_{i \in I} \in \prod M_i, \ (\prod_{i \in I} v_i)(x) = (v_i(x_i))_{i \in I} \in \prod N_i. \)

Similarly one defines the **direct sum** of the family of homomorphisms \( (v_i)_{i \in I}. \) Let \( (\oplus M_i, (\sigma_i)_{i \in I}) \) (respectively \( (\oplus N_i, (\tau_i)_{i \in I}) \) be the direct sum of the family \( (M_i)_{i \in I} \) (respectively \( (N_i)_{i \in I} \)). For any \( j \in I, \)

\( \tau_j \circ v_j : M_j \to \oplus N_i \) is a homomorphism. The universality property of the
direct sum \((3.5)\) yields a unique homomorphism \(w : \bigoplus M_i \to \bigoplus N_i\) such that \(w \circ \sigma_j = \tau_j \circ v_j, \forall j \in I\).

\[
\begin{array}{ccc}
M_j & \xrightarrow{\sigma_j} & \bigoplus M_i \\
\downarrow{v_j} & & \downarrow{w} \\
N_j & \xrightarrow{\tau_j} & \bigoplus N_i
\end{array}
\]

The homomorphism \(w\) is called the direct sum of the family \((v_i)_{i \in I}\) and is denoted by \(\prod_{i \in I} v_i\) or \(\bigoplus_{i \in I} v_i\). If \(I = \{1, \ldots, n\}\), \(w\) is denoted by \(\bigoplus^n_{i=1} v_i\) or \(v_1 \oplus \ldots \oplus v_n\).

If \(x = \sum_{i \in I} x_i \in \bigoplus M_i\), where \(x_i \in M_i\) and the family \((x_i)_{i \in I}\) has finite support\(^{11}\), then \((\bigoplus_{i \in I} v_i)(x) = \sum_{i \in I} v_i(x_i) \in \bigoplus N_i\).

**3.18 Proposition.** Let \(v_i : M_i \to N_i\) \((i \in I)\) be a family of module homomorphisms. Then:

a) If \(v_i\) is injective (surjective), then \(\prod_{i \in I} v_i\) and \(\bigoplus_{i \in I} v_i\) are injective (surjective).

b) If \(u_i : N_i \to P_i\) \((i \in I)\) are homomorphisms, then \((\prod_{i \in I} u_i)^v(\prod_{i \in I} v_i) = \prod_{i \in I} u_i^v v_i\) and \((\bigoplus_{i \in I} u_i)^v(\bigoplus_{i \in I} v_i) = \bigoplus_{i \in I} u_i^v v_i\). \(\square\)

Let \(R\text{-Mod}\) be the category of the left \(R\)-modules and \(Ab\) the category of Abelian groups.

**3.19 Definition.** (The Hom functors) For any \(A \in R\text{-Mod}\), define the (covariant) functor \(h^A : R\text{-Mod} \to Ab\):

\[
\forall B \in R\text{-Mod}, h^A(B) := \text{Hom}_R(A, B),
\]

(note that \(\text{Hom}_R(A, B)\) is an Abelian group with respect to homomorphism addition);

---

\(^{11}\) We identify \(x_i\) with its image through the canonical injection \(\sigma_i(x_i)\).
∀ u : B → B’ morphism in R-Mod, h^A(u) : Hom_R(A, B) → Hom_R(A, B’)

is defined as

h^A(u)(g) := u \circ g, \forall g \in Hom_R(A, B).

It is immediate that h^A(u) is a morphism in Ab, that h^A(1_B) = 1_{h^A(B)}
and that h^A(v \circ u) = h^A(v) \circ h^A(u), for any R-module B and any R-module
homomorphisms u : B → B’ and ∀ v : B’ → B”. So, h^A is a functor, also
denoted by Hom_R(A, -).

In a similar manner one defines the contravariant functor
h_A : R-Mod → Ab. For any B ∈ R-Mod, h_A(B) := Hom_R(B, A); for any
u : B → B’ in R-Mod, h_A(u) : Hom_R(B’, A) → Hom_R(B, A) is given by
h_A(u)(g) := g \circ u, \forall g \in Hom_R(B’, A). The functor h_A is denoted also by
Hom_R(-, A).

We study the behavior of Hom_R(A, -) and Hom_R(-, A) with respect
to direct products and direct sums.

**3.20 Proposition.** Let A be an R-module and let (M_i)_{i \in I} be a family
of R-modules. Then there exist canonical isomorphisms in Ab:

Hom_R(A, \prod_{i \in I} M_i) \cong \prod_{i \in I} Hom_R(A, M_i),

Hom_R(\bigoplus_{i \in I} M_i, A) \cong \prod_{i \in I} Hom_R(M_i, A).

**Proof.** We prove the second isomorphism (the first is proposed as an exercise).

Let S = \bigoplus_{i \in I} M_i and let \sigma_i : M_i → S be the canonical injections.

We give two proofs, a “direct proof” and a “category proof”. The
“direct proof” uses the form of the elements in the concrete constructions
of the direct sum and the direct product and defines a natural
homomorphism from Hom_R(S, A) to \prod_{i \in I} Hom_R(M_i, A); we then show
this is an isomorphism. The “category proof” shows that Hom_R(S, A) is
a direct product in Ab of the family (Hom_R(M_i, A))_{i \in I} (more precisely,
it satisfies the universality property of the direct product) and applies
then 3.3.

**Direct proof.** Define \alpha : Hom_R(S, A) → \prod_{i \in I} Hom_R(M_i, A) as fol-
llows: for any homomorphism \( g : \bigoplus_{i \in I} M_i \to A \), \( \alpha(g) =: (g \circ \sigma_i)_{i \in I} \in \prod_{i \in I} \text{Hom}_R(M_i, A) \).

It is routine to check that \( \alpha \) is a homomorphism. If \( \alpha(g) = (g \circ \sigma_i)_{i \in I} = (0)_{i \in I} \), then \( g(\sum_{i \in I} \sigma_i(x_i)) = \sum_{i \in I} g(\sigma_i(x_i)) = 0 \), for any family having finite support \((x_i)_{i \in I}\), with \( x_i \in M_i \), so \( g = 0 \). This proves that \( \alpha \) is injective. If \((g_i)_{i \in I} \in \prod_{i \in I} \text{Hom}_R(M_i, A)\), the universality property of the direct sum \( S = \bigoplus_{i \in I} M_i \) (definition 3.5) supplies a unique homomorphism \( g : S \to A \) such that \( g \circ \sigma_i = g_i, \forall i \in I \). So, \( \alpha \) is surjective.

**Category proof:** For any \( i \in I \), \( h_A(\sigma_i) : h_A(S) \to h_A(M_i) \) is a morphism in \( Ab \). We show that \( h_A(S) \) is a direct product of the Abelian groups \( \{h_A(M_i)\}_{i \in I} \), with canonical projections \( h_A(\sigma_i) \). This means that we have to prove that, for any \( X \in Ab \) and any morphisms \( \tau_i : X \to h_A(M_i) \) in \( Ab \), \( i \in I \), there exists a unique morphism \( \tau : X \to h_A(S) \) in \( Ab \) such that \( \tau_i = h_A(\sigma_i) \circ \tau \). So, \( \forall x \in X, \; \tau_i(x) : M_i \to A \) is a morphism in \( R-\text{Mod} \). By the universality property of the direct sum, \( \exists! \; \tau(x) : S \to A \) homomorphism such that \( \tau_i(x) = \tau(x) \circ \sigma_i \). We obtain \( \tau : X \to \text{Hom}_R(S, A) = h_A(S), \; x \mapsto \tau(x) \), which is a morphism in \( Ab \). Indeed, \( \forall x, \; y \in X, \; \tau(x) + \tau(y) : S \to A \) is an \( R \)-homomorphism satisfying

\[
(\tau(x) + \tau(y)) \circ \sigma_i = \tau(x) \circ \sigma_i + \tau(y) \circ \sigma_i = \tau_i(x) + \tau_i(y) = \tau(x + y).
\]

Since \( \tau(x + y) \) is the only homomorphism with this property, \( \tau(x + y) = \tau(x) + \tau(y) \). If \( \varphi : X \to h_A(S) \) is a homomorphism with \( \tau_i = h_A(\sigma_i) \circ \varphi \), then, \( \forall x \in X, \; \varphi(x) : S \to A \) is a homomorphism with \( \tau_i(x) = h_A(\sigma_i)(\varphi(x)) = \varphi(x) \circ \sigma_i \). But \( \tau(x) \) is unique with this property, so \( \tau(x) = \varphi(x) \), \( \forall x \in X \), i.e. \( \tau = \varphi \). \( \square \)

Many facts in Algebra have a convenient form in the language of **exact sequences** (used intensively Homological Algebra, for instance).
3.21 Definition. Consider a (finite or infinite) sequence of $R$-modules and module homomorphisms:\(^{12}\)

(S): \[ \ldots \longrightarrow E \xrightarrow{u} F \xrightarrow{v} G \longrightarrow \ldots, \]

The sequence (S) is called semiexact at $F$ if $v \circ u = 0$. This may be rephrased as $\text{Im } u \subseteq \text{Ker } v$. The sequence (S) is called exact at $F$ if $\text{Im } u = \text{Ker } v$. The sequence (S) is called semiexact (respectively exact) if it is semiexact (respectively exact) at any term. A semiexact sequence of modules is called a complex of modules.

3.22 Remarks. a) Often, the homomorphisms that are uniquely determined (or clear from the context) are not marked on the arrows. For instance, write $0 \rightarrow L$ instead of $0 \xrightarrow{0} L$, since the only homomorphism defined on the module 0 is the zero homomorphism.

b) A sequence of the form $0 \rightarrow E \xrightarrow{u} F$ is exact if and only if $u$ is a monomorphism. Indeed, $\text{Ker } u = 0 \iff \text{Ker } u = \text{Im } 0$.

c) A sequence of the form $E \xrightarrow{u} F \rightarrow 0$ is exact if and only if $u$ is an epimorphism.

d) The sequence $0 \rightarrow E \xrightarrow{u} F \rightarrow 0$ is exact $\iff u$ is an isomorphism.

3.23 Examples. a) If $\varphi : E \rightarrow F$ is a homomorphism, then the following sequence is exact:

\[ 0 \rightarrow \text{Ker } \varphi \xrightarrow{\iota} E \xrightarrow{\varphi} F \xrightarrow{\pi} F/\text{Im } \varphi \rightarrow 0 \]

The inclusion homomorphism is denoted by $\iota$ and the canonical surjection by $\pi$. The module $F/\text{Im } \varphi$ is called the cokernel of $\varphi$ and is denoted by $\text{Coker } \varphi$. Thus, we have an exact sequence:

\[ 0 \rightarrow \text{Ker } \varphi \xrightarrow{\iota} E \xrightarrow{\varphi} F \xrightarrow{\pi} \text{Coker } \varphi \rightarrow 0. \]

b) If $A \leq_R B$, then the sequence

\(^{12}\) Called in the sequel sequence of modules.
0 \to A \xrightarrow{t_A} A \oplus C \xrightarrow{\pi_C} C \to 0 \text{ is short exact; } t_A : A \to A \oplus C \text{ is the canonical injection and } 
\pi_C : A \oplus C \to C \text{ is the canonical projection (recall that } A \oplus C = A \times C). \text{ So } A \oplus C \text{ is an extension of } A \text{ by } C. \text{ The problem of finding all extensions of } A \text{ by } C \text{ is highly nontrivial. It is natural in this sense to consider the following definition:}

3.26 Definition. Let } A \text{ and } C \text{ be two modules and consider two extensions of } A \text{ by } C:

\[
0 \to A \xrightarrow{u} B \xrightarrow{v} C \to 0, \\
0 \to A \xrightarrow{u'} B' \xrightarrow{v'} C \to 0.
\]

We say that the extensions are equivalent if there exists a homomorphism } g : B \to B' \text{ such that } g \circ u = u' \text{ and } v' \circ g = v. \text{ In other words, the diagram below is commutative:}

\[
\begin{array}{ccc}
0 & \to & A \\
\downarrow & & \downarrow^g \\
0 & \to & A
\end{array}
\begin{array}{ccc}
\to & B & \to C \\
\downarrow & & \downarrow \\
\to & B' & \to C
\end{array}
\]

(\text{the vertical unmarked homomorphisms are identity homomorphisms}). A homomorphism } g \text{ that makes the diagram commutative is called a homomorphism of extensions.}
3.27 Proposition. In the definition 3.26, $g$ is an isomorphism. In particular, the relation defined at 3.26 is an equivalence relation on the class of all extensions of $A$ by $C$.

Proof.\textsuperscript{13} Let us show that $\ker g = 0$. Let $b \in B$ with $g(b) = 0$. Then $v'g(b) = 0 = v(b)$, so $b \in \ker v = \operatorname{Im} u$. Thus there exists $a \in A$ with $u(a) = b$; we obtain $0 = g(b) = gu(a) = u'(a)$; since $u'$ is injective, $a = 0$, so $b = u(a) = 0$.

Let $b' \in B'$; then $v'(b') \in C$ and the surjectivity of $v$ implies that, for some $b \in B$, $v(b) = v'(b')$. Since $v'g(b) = v(b)$, $v'(g(b)) = v'(b')$, which means $g(b) - b' \in \ker v' = \operatorname{Im} u'$. There exists $a \in A$ with $u'(a) = g(b) - b'$; since $u' = gu$, $g(b) - b' = u'(a) = gu(a)$. Thus $b' = g(b - u(a)) \in \operatorname{Im} g$. \hfill \square

3.28 Definition. Consider two complexes:

$E$: \[ \cdots \rightarrow E_{i-1} \xrightarrow{u_{i-1}} E_i \xrightarrow{u_i} E_{i+1} \rightarrow \cdots \]

$F$: \[ \cdots \rightarrow F_{i-1} \xrightarrow{v_{i-1}} F_i \xrightarrow{v_i} F_{i+1} \rightarrow \cdots \]

A homomorphism of complexes from $E$ to $F$ is a sequence of module homomorphisms $g := (g_i)_{i \in \mathbb{Z}}$, $g_i : E_i \rightarrow F_i$, such that $g_{i+1} \circ u_i = v_i \circ g_i$, $\forall i \in \mathbb{Z}$. In other words, the following diagram commutes:

\[
\begin{array}{ccc}
\cdots & \xrightarrow{u_{i-1}} & E_i & \xrightarrow{u_i} & E_{i+1} & \rightarrow \cdots \\
\downarrow g_{i-1} & & \downarrow g_i & & \downarrow g_{i+1} & \\
\cdots & \xrightarrow{v_{i-1}} & F_i & \xrightarrow{v_i} & F_{i+1} & \rightarrow \cdots 
\end{array}
\]

The homomorphism $g$ is called an isomorphism if $g_i$ is an isomorphism, $\forall i \in \mathbb{Z}$. In this case, $(g_i^{-1})_{i \in \mathbb{Z}}$ is also a homomorphism of complexes. The complexes $E$ and $F$ are called isomorphic if there exists an isomorphism from $E$ to $F$.

\textsuperscript{13} The technique used is called “diagram chasing” and it is used extensively in arguments involving diagrams of module homomorphisms.
3.29 Examples. a) The exact sequence $0 \rightarrow A \overset{u}{\rightarrow} B \overset{v}{\rightarrow} C \rightarrow 0$ is isomorphic (in the sense given by the previous definition) to the sequence $0 \rightarrow \text{Im} u \overset{t}{\rightarrow} B \overset{\pi}{\rightarrow} B/\text{Im} u \rightarrow 0$, where $t$ is the inclusion and $\pi$ the canonical surjection. The homomorphisms that compose this isomorphism are $u_0 : A \rightarrow \text{Im} u$, $u_0(a) = u(a)$, $\forall a \in A$; $\text{id}_B : B \rightarrow B$; in order to define $w : C \rightarrow B/\text{Im} u$, observe that the surjective homomorphism $v$ induces a canonical isomorphism $v_0 : B/\text{Ker} v \rightarrow C$, $v_0(b + \text{Ker} v) = v(b)$, $\forall b \in B$ (cf. the fundamental isomorphism theorem). Set $w = v_0^{-1}$. Checking the commutativity of the diagram is left to the reader.

b) Any extension of $\mathbb{Z}_2$ by $\mathbb{Z}_2$ is isomorphic either to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ or to $\mathbb{Z}_4$.

The extension “direct sum” of $A$ by $C$,

$$0 \rightarrow A \overset{t_A}{\rightarrow} A \oplus C \overset{\pi_C}{\rightarrow} C \rightarrow 0,$$

has the remarkable property that by reversing the arrows we obtain another exact sequence: $0 \leftarrow A \leftarrow \pi_A A \oplus C \leftarrow i_C C \leftarrow 0$. The extensions of $A$ by $C$ that are isomorphic to the extension direct sum $A \oplus C$ are characterized in the next proposition:

3.30. Proposition. Let $(S) : 0 \rightarrow A \overset{u}{\rightarrow} B \overset{v}{\rightarrow} C \rightarrow 0$ be a short exact sequence of modules (i.e., $B$ is an extension of $A$ by $C$). The following statements are equivalent:

a) $B$ is isomorphic (as an extension of $A$ by $C$) with $0 \rightarrow A \overset{t_A}{\rightarrow} A \oplus C \overset{\pi_C}{\rightarrow} C \rightarrow 0$, the “direct sum” extension, where $t_A, i_C$ are the canonical injections of the direct sum $A \oplus C$ and $\pi_A, \pi_C$ are the canonical projections.

b) $\text{Im} u = \text{Ker} v$ is a direct summand in $B$.

c) There exists a homomorphism $u' : B \rightarrow A$ such that $u' \circ u = \text{id}_A$.

d) There exists a homomorphism $v' : C \rightarrow B$ such that $v \circ v' = \text{id}_C$.

Proof. c)$\Rightarrow$b) and d)$\Rightarrow$b) follow from 3.16.
II. Modules

a)⇒c), d) Let \( \varphi : B \to A \oplus C \) be an isomorphism such that the diagram

\[
\begin{array}{ccc}
0 & \to & A \xrightarrow{u} B \xrightarrow{v} C \to 0 \\
\downarrow & & \downarrow \varphi \\
0 & \to & A \oplus C \xrightarrow{\pi_C} C \to 0
\end{array}
\]

commutes. Define \( u' : B \to A, \quad u' = \pi_A \varphi \). We have \( u'u = \pi_A \varphi u = \pi_A \iota_A = \text{id}_A \) (see Remark 3.14). This proves c). Defining \( v' : C \to B, \quad v' = \varphi^{-1} \iota_C \), we have \( vv' = v \varphi^{-1} \iota_C = \pi_C \iota_C = \text{id}_C \), so d) holds.

b)⇒a) Let \( B = \text{Im} \ u \oplus D \); thus, \( \forall b \in B, \ \exists! \ a \in A \) and \( d \in D \) such that \( b = u(a) + d \). Define then \( \alpha(b) = a \in A \). We obtain a module homomorphism \( \alpha : B \to A \). Indeed, if \( b = u(a) + d, \ b' = u(a') + d' \in B \), with \( a, a' \in A, \ d, d' \in D \), then

\[ b + b' = u(a) + d + u(a') + d' = u(a + a') + d + d'; \]

so, \( \alpha(b + b') = a + a' = \alpha(b) + \alpha(b') \).

In the same way is shown that \( \alpha \) preserves multiplication with scalars. Note that \( \alpha \circ u : A \to A \) is \( \text{id}_A \).

Consider \( \varphi : B \to A \oplus C, \ \varphi = \iota_A \alpha + \iota_C v \). We have

\[
\varphi u = (\iota_A \alpha + \iota_C v) u = \iota_A au + \iota_C vu = \iota_A
\]

\[
\pi_C \varphi = \pi_C (\iota_A \alpha + \iota_C v) = \pi_C \iota_A \alpha + \pi_C \iota_C v = v.
\]

So, \( \varphi \) is a homomorphism (and isomorphism) of extensions. \( \square \)

3.31 Definition. A short exact sequence

\[
0 \to A \xrightarrow{u} B \xrightarrow{v} C \to 0
\]

is called split (or we say, “the sequence splits”) if it satisfies the equivalent conditions of the previous proposition.

A monomorphism \( 0 \to A \xrightarrow{u} B \) is called split if the short exact sequence \( 0 \to A \xrightarrow{u} B \to B/\text{Im} u \to 0 \) splits (\( \iff \text{Im} u \) is a direct summand in \( B \)).

An epimorphism \( B \xrightarrow{v} C \to 0 \) is called split if the short exact sequence \( 0 \to \text{Ker} v \to B \xrightarrow{v} C \to 0 \) splits (\( \iff \text{Ker} v \) is a direct summand in \( B \)).
The behavior with respect to short exact sequences is primordial in the study of functors defined on categories of modules. The next proposition says “the functor Hom is left exact”:

3.32 Proposition. Let $R M$ be a module and let

\[ 0 \to A \xrightarrow{u} B \xrightarrow{v} C \to 0 \]

be a short exact sequence. Then the following sequence

\[ 0 \to h^M(A) \xrightarrow{h^M(u)} h^M(B) \xrightarrow{h^M(v)} h^M(C) \]

is exact in $\text{Ab}$. In particular, if $u : A \to B$ is an $R$-module monomorphism, then $h^M(u)$ is an Abelian group monomorphism.

**Proof.** We have to prove that $h^M(u)$ is monomorphism and $\text{Im } h^M(u) = \text{Ker } h^M(v)$. Let $\varphi \in h^M(A)$ (i.e. $\varphi : M \to A$ is a homomorphism) such that $h^M(u)(\varphi) = u \circ \varphi = 0$. Since $u$ is monomorphism, $\varphi = 0$. So, $\text{Ker } h^M(u) = 0$.

Since $h^M(v) \circ h^M(u) = h^M(v \circ u) = h^M(0) = 0$, $\text{Im } h^M(u) \subseteq \text{Ker } h^M(v)$. The reverse inclusion also holds: if $\psi \in h^M(B)$ (i.e. $\psi : M \to B$) is in $\text{Ker } h^M(v)$, then $v \circ \psi = 0$, so $\text{Im } \psi \subseteq \text{Ker } v = \text{Im } u$. But $u$ is injective, so $u_0 : A \to \text{Im } u \quad (u_0(a) = u(a), \quad \forall a \in A)$ is an isomorphism; let $u' : \text{Im } u \to A$ be the inverse of $u_0$. Then $u' \circ \psi : M \to B$ is well defined and it is a homomorphism. Also, $h^M(u)(u' \circ \psi) = u \circ u' \circ \psi = \psi$.

If $u$ is monomorphism, consider the exact sequence:

\[ 0 \to A \xrightarrow{u} B \to B/\text{Im } u \to 0 \]

and write that $h^M$ is left exact to obtain that $h^M(u)$ is monomorphism. □

The contravariant functor $h_M$ is also left exact. We invite the reader to prove:

3.33 Proposition. Let $R M$ be an $R$-module and let

\[ 0 \to A \xrightarrow{u} B \xrightarrow{v} C \to 0 \]

be a short exact sequence. Then the sequence

\[ 0 \to h_M(C) \xrightarrow{h_M(v)} h_M(B) \xrightarrow{h_M(u)} h_M(A) \]
is exact in $\text{Ab}$. In particular, if $\nu : B \to C$ is an $R$-module epimorphism, then $h_M(\nu)$ is a monomorphism.

Exercises

In the exercises, $R$ denotes a fixed ring with identity.

1. Prove that the $R$-module $S$ is a direct sum of the family of $R$-modules $(M_i)_{i \in I}$, with canonical injections $(\sigma_i)_{i \in I}$, if and only if for any $x \in S$, there exists a unique family of elements $(x_i)_{i \in I}$, with $x_i \in M_i$, $\forall i \in I$, having finite support, such that $x = \sum_{i \in I} \sigma_i(x_i)$.

2. Let $A$ and $B$ be $R$-modules. Prove that there exists a (canonical) isomorphism $A \oplus B \cong B \oplus A$. Generalization.

3. Prove that the $R$-module $P$ is a direct product of family of $R$-modules $(M_i)_{i \in I}$, with canonical projections $(\pi_i)_{i \in I}$, if and only if for any family $(x_i)_{i \in I}$, with $x_i \in M_i$, $\forall i \in I$, there exists a unique $x \in P$ such that $\pi_i(x) = x_i$, $\forall i \in I$.

4. Let $R$ be a ring and let $I$ be a left ideal of $R$. For any module $_RA$, define the submodule of $A$:

$$IA := \langle \{ia \mid i \in I, a \in A\} \rangle.$$ 

a) Prove that $IA = \{i_1a_1 + \ldots + i_na_n \mid n \in \mathbb{N}, \ i_1, \ldots, i_n \in I, \ a_1, \ldots, a_n \in A\}.$

b) If $A = \bigoplus_{s \in S} A_s$ (internal direct sum), then $IA = \bigoplus_{s \in S} IA_s$.

c) If $A = \prod_{s \in S} A_s$, then $IA \subseteq \prod_{s \in S} IA_s$ (we identify $A_s$ with its image in $\prod_{s \in S} A_s$; the same applies to $IA_s$). If $I$ is finitely generated, then the inclusion is in fact an equality. Prove that the inclusion may be strict.

5. Taking the universality property of the direct product as its definition, prove that the canonical projections are surjective.
6. Find homomorphisms $u$, $v$, $w$ such that the following sequence is exact:

$$0 \to \mathbb{Z}/3 \xrightarrow{u} \mathbb{Z}/9 \xrightarrow{v} \mathbb{Z}/9 \xrightarrow{w} \mathbb{Z}/3 \to 0.$$ 

7. Let $R$ be a PID. Prove that the $R$-module $M$ is cyclic (generated by one element) $\iff$ there exists an exact sequence of the form $0 \to R \to R \to M \to 0$.

8. Let $R$ be a ring and let $M$ be an $R$-module. Then any $R$-epimorphism $\varphi : M \to R$ splits. Is it true that for any ring $R$ and for any $R$-module $M$, any $R$-monomorphism $\psi : R \to M$ splits?

9. Suppose that the following diagram of $R$-modules commutes and has exact rows:

$$\xymatrix{ A \ar[r]^u \ar[d]^{\alpha} & B \ar[r]^v \ar[d]^{\beta} & C \ar[d]^{\gamma} \\
A' \ar[r]^{u'} & B' \ar[r]^{v'} & C' }$$

Prove, using diagram chasing, that:

a) If $\alpha$, $\gamma$ and $u'$ are monomorphisms, then $\beta$ is monomorphism.

b) If $\alpha$, $\gamma$ and $v$ are epimorphisms, then $\beta$ is an epimorphism.

c) If $\beta$ is a monomorphism and $\alpha$, $v$ are epimorphisms, then $\gamma$ is an epimorphism.

d) If $\beta$ is an epimorphism and $u'$, $\gamma$ are monomorphisms, then $\alpha$ is a monomorphism.

10. Give an example of an $R$-module $M$ with a direct summand $S$ in $M$ that has an infinity of complements.

11. Show that any composition of split monomorphisms (epimorphisms) is a split monomorphism (epimorphism). Give an example of monomorphisms $u$, $v$ such that $uv$ splits, but $v$ does not split.
II. Modules

II.4 Free modules

Many nice properties of the vector spaces are consequences of the fact that any vector space has a basis. The existence of a basis is not guaranteed for a module, and this increases considerably the difficulty in the study of modules in comparison to the vector spaces. It is natural to define and to study the class of modules that “have a basis”.

In this section, $R$ is a ring with identity and all modules are left $R$-modules.

4.1. Definition. Let $M$ be an $R$-module, $n \in \mathbb{N}^*$ and let $x_1, \ldots, x_n$ be a family of elements of $M$. We say that the family $x_1, \ldots, x_n$ is linearly independent (over $R$) \footnote{The reference to the ring $R$ is often omitted.} if, for any $r_1, \ldots, r_n \in R$, $r_1x_1 + \ldots + r_nx_n = 0$ implies $r_1 = \ldots = r_n = 0$. In other words, any linear combination of $x_1, \ldots, x_n$ is 0 if and only if all its coefficients are 0.

A family of elements in $M$ that is not linearly independent is called linearly dependent. A relation of the form $r_1x_1 + \ldots + r_nx_n = 0$, with $r_1, \ldots, r_n \in R$, not all zero, is called a relation of linear dependence of the family $x_1, \ldots, x_n$.

4.2. Remarks. a) If there exists $i \neq j$ with $x_i = x_j$, then the family $x_1, \ldots, x_n$ is linearly dependent: the linear combination $x_i - x_j$ is 0. Thus, in studying linear independence we may suppose that $x_1, \ldots, x_n$ are distinct. On the other hand, the notion of linear independence does not depend on the indexing of $x_1, \ldots, x_n$. This is the reason we can speak about a linearly dependent (finite) subset of $M$.

b) The set $\{x\}$ (containing a single element $x \in M$) is linear independent if and only if $\forall r \in R$, $rx = 0$ implies $r = 0$. This suggest the following definition: the annihilator of $x$ in $R$ is
Ann\(_R(x) := \{ r \in R \mid rx = 0 \} \) (sometimes denoted by \( l_R(x) \)). Since Ann\(_R(x)\) is exactly the kernel of the \( R\)-module homomorphism \( \rho_x : R \to M, \rho_x(r) = rx, \ \forall r \in R \), Ann\(_R(x)\) is a left ideal in \( R \) and (by the isomorphism theorem) the submodule \( Rx = \text{Im} \rho_x \) is isomorphic as a left \( R\)-module to \( R/\text{Ann}_R(x) \). Summarizing: \( \{x\} \) is linearly independent \( \iff \text{Ann}_R(x) = 0 \iff \rho_x \) is an isomorphism from \( R \) to \( Rx \).

4.3. Definition. Let \( X \) be a subset of \( R \cdot M \). We say that \( X \) is linearly independent over \( R \) (or free over \( R \)) if any finite subset of \( X \) is linearly independent over \( R \) in the sense of the remark above. If \( X \) is not linearly independent, we say that \( X \) is linearly dependent over \( R \). Thus, \( X \) is linearly dependent if and only if there exists a finite linearly dependent subset of \( X \).

4.4. Remarks.  
a) The empty subset of \( M \) is linearly independent. 
b) Let \( \{x_i\}_{i \in I} \) be a family of elements of the \( R\)-module \( M \). The family \( \{x_i\}_{i \in I} \) is linearly independent if and only if, for any family \( \{r_i\}_{i \in I} \) of elements of \( R \), with finite support, \( \sum_{i \in I} r_ix_i = 0 \) implies \( r_i = 0, \ \forall i \in I \).

It follows that the sum of the family of submodules \( \{Rx_i\}_{i \in I} \) is direct (see 3.11). In fact, \( \{x_i\}_{i \in I} \) is linearly independent if and only if the sum of the family of submodules \( \{Rx_i\}_{i \in I} \) is direct and \( \text{Ann}_R(x_i) = 0, \forall i \in I \) (prove this!).

c) Every subset of an linearly independent set is linearly independent.

4.5. Examples.  
a) The set \( \{1\} \), containing only the unity of the ring \( R \) (seen as a left \( R\)-module) is linearly independent. More generally, \( \forall r \in R, \ \{r\} \) is linearly dependent \( \iff r \) is a right zero divisor in \( R \) (\( \exists s \in R, s \neq 0, \text{ such that } sr = 0 \)).
b) The $\mathbb{Z}$-module $\mathbb{Z}_3$ has no $\mathbb{Z}$-linearly independent subsets: for any $x \in \mathbb{Z}_3$, $3x = \hat{0}$. Can you generalize this? Of course, the set $\{\hat{1}\}$ is linearly independent over $\mathbb{Z}_3$, as seen at example a).

c) If $R$ is a domain, in the $R$-module $R[X]$ the set $\{X^n \mid n \in \mathbb{N}\}$ is linearly independent. This amounts to saying that a polynomial $a_0 + a_1X + ... + a_nX^n$ $(a_0, a_1, ..., a_n \in R)$ is 0 if and only if $a_0 = ... = a_n = 0$. More generally, if $\{f_n \mid n \in \mathbb{N}\}$ is a family of polynomials such that $\text{deg} f_n \neq \text{deg} f_m$ if $m \neq n$, then the family $\{f_n \mid n \in \mathbb{N}\}$ is linearly independent.

4.6. Definition. A subset $B$ of an $R$-module $M$ is called a basis of $M$ if it is simultaneously linearly independent and a system of generators for $M$. The module $M$ is called a free module if it has a basis.

4.7 Proposition. Let $M$ be an $R$-module and let $B$ be a subset of $M$. Then: $B$ is a basis of $M$ if and only if any $x \in M$ is written uniquely as a linear combination of elements of $B$.

Proof. Let $B = \{e_i \mid i \in I\}$ be a basis of $R M$ and let $x \in M$. Since $B$ generates $M$, there exists a family $\{r_i \mid i \in I\}$ of elements in $R$ with finite support, such that $x = \sum_{i \in I} r_i e_i$. If $\{s_i \mid i \in I\}$ is another family of elements in $R$, with $x = \sum_{i \in I} s_i e_i$, then $\sum_{i \in I} (s_i - r_i) e_i = 0$. The linear independence of $B$ implies $s_i = r_i$, $\forall i \in I$.

Conversely, if any element in $M$ is written uniquely as a linear combination of elements of $B$, then $B$ is a system of generators of $M$. Writing the fact that 0 has a unique writing as a linear combination of elements of $B$, we obtain that $B$ is linearly independent. \]

With the notations above, for $x \in M$, the elements $\{r_i \mid i \in I\}$ in $R$ with $x = \sum_{i \in I} r_i e_i$ are called the coordinates of $x$ in the basis $B = \{e_i \mid i \in I\}$.

4.8. Examples. a) $\{\emptyset\}$ is a basis (the only one!) of the module $\{0\}$.

b) $\{1\}$ is a basis of $R R$. More generally, $\{r\}$ is a basis of $R R$ if and only if $r$ is right invertible in $R$. 
II.4 Free modules

c) If $I$ is a set, the $R$-module $R^{(I)}$ (the direct sum of $|I|$ copies of $R$, with canonical injections $\sigma_i : R \to R^{(I)}$) is free, a basis being $\{e_i\}_{i \in I}$, where $e_i = \sigma_i(1)$. This basis is called the canonical basis of $R^{(I)}$; $R^{(I)}$ is also called the free $R$-module on the set $I$ (or the free $R$-module of basis $I$). If $I = \{1, \ldots, n\}$, the canonical basis of the free $R$-module $R^n$ are $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $\ldots$, $e_n = (0, 0, \ldots, 1)$.

A module homomorphism is determined by its values on a generating set. But, for an arbitrary generating set, there may be no homomorphism that takes prescribed values on the elements of the generating set. In the privileged case of free modules, for any choice of the values on the elements of a basis, a unique homomorphism takes the respective values on the elements of the basis:

4.9 Proposition. (The universality property of the free module of basis $B$) Let $L$ be a free $R$-module of basis $B$ and let $i : B \to L$ be the inclusion map. The pair $(L, i)$ has the following universality property:

For any pair $(M, u)$, where $M$ is an $R$-module and $u : B \to M$ is a map, there exists a unique homomorphism $v : L \to M$ with $v \circ i = u$.

If $B = (e_i)_{i \in I}$, this can be formulated:

For any module $M$ and any family $(y_i)_{i \in I}$ of elements of $M$, there exists a unique homomorphism $v : L \to M$ such that $v(e_i) = y_i$, $\forall i \in I$.

Moreover, we have:

a) $v$ is injective $\iff u(B)$ is linearly independent.

b) $v$ is surjective $\iff u(B)$ is a system of generators for $M$.

c) $v$ is an isomorphism $\iff u(B)$ is a basis.

Proof. Uniqueness claim: suppose $v, w \in \text{Hom}_R(L, M)$ satisfy $v(e_i) = y_i = w(e_i)$, $\forall i \in I$. Let $x \in L$; since $(e_i)_{i \in I}$ is a basis, there exists a unique family $(r_i)_{i \in I}$ of elements in $R$ such that $x = \sum_{i \in I} r_i e_i$. We have $v(\sum_{i \in I} r_i e_i) = \sum_{i \in I} r_i v(e_i) = \sum_{i \in I} r_i y_i$. The same thing is obtained for $w$.

For the existence claim, we have to prove that
II. Modules

\( v(x) := \sum_{i \in I} r_i y_i \), for any \( x = \sum_{i \in I} r_i e_i \) with \( (r_i)_{i \in I} \) family of elements in \( R \), having a finite support,

defines a module homomorphism. Since, for any \( x \in L \) there exists a unique family \( (r_i)_{i \in I} \) of elements in \( R \) such that \( x = \sum_{i \in I} r_i e_i \), \( v \) is well defined. Let \( x = \sum_{i \in I} r_i e_i \) and \( y = \sum_{i \in I} s_i e_i \) be elements in \( L \), with \( (r_i)_{i \in I} \), \( (s_i)_{i \in I} \) families of elements in \( R \). For any \( a, b \in R \), we have:

\[
v(ax + by) = v(\sum_{i \in I} ar_i e_i + \sum_{i \in I} bs_i e_i) = v(\sum_{i \in I} (ar_i + bs_i)e_i) = \sum_{i \in I} (ar_i + bs_i)y_i = av(x) + bv(y).
\]

We prove \( a) \): \( \ker v = \{ \sum_{i \in I} r_i e_i \mid (r_i)_{i \in I} \in R^I, \sum_{i \in I} r_i y_i = 0 \} \). It is clear that \( \ker v = 0 \iff (y_i)_{i \in I} \) is linearly independent.

The rest of the proof is left to the reader. \( \square \)

The \( R \)-modules of the type \( R^I \) are “all” free \( R \)-modules:

**4.10 Proposition.** Let \( L \) be a free \( R \)-module and let \( (x_i)_{i \in I} \) be a basis of \( L \). Then \( R^I \cong R L \).

**Proof.** Let \( (e_i)_{i \in I} \) be the canonical basis in \( R^I \). By the result above, there exists a unique homomorphism \( u : L \to R^I \) with \( u(x_i) := e_i \), \( \forall i \in I \). Since \( \text{Im} \ u \) includes \( \{ e_i \}_{i \in I} \), which generates \( R^I \), it follows that \( u \) is surjective. If \( x = \sum_{i \in I} r_i x_i \in \ker u \) (with \( (r_i)_{i \in I} \) family of elements in \( R \)), then \( 0 = u(x) = \sum_{i \in I} r_i e_i \). Since \( (e_i)_{i \in I} \) is a basis, \( r_i = 0 \), \( \forall i \in I \), so \( x = 0 \). \( \square \)

A direct sum of free modules is also free:

**4.11 Proposition.** Let \( (M_i)_{i \in I} \) be a family of free \( R \)-modules. Then the (external) direct sum \( \bigoplus_{i \in I} M_i \) is a free \( R \)-module.

**Proof.** Let \( M := \bigoplus_{i \in I} M_i \). For any \( j \in I \), let \( \sigma_j : M_j \to M \) be the canonical injection and let \( B_j \) be a basis of \( M_j \). We show that \( B := \bigcup_{i \in I} \sigma_j(B_i) \) is a basis of \( M \). By 3.10, \( M \) is the (internal) direct sum
of the family of submodules \((\text{Im } \sigma_i)_{i \in I}\). Since \(\sigma_i(B_i)\) generates \(\text{Im } \sigma_i\), \(\forall i \in I, \bigcup_{i \in I} \sigma_i(B_i)\) generates \(M\). The linear independence of \(B\) is easy to show (only the notations are complicated): let \(B_i = (e_{it})_{t \in T_i}\) the basis in \(M_i\) and let \(y_{it} = \sigma_i(e_{it})\). If \(\{r_{it} | i \in I, t \in T_i\}\) is a finite support family of elements of \(R\), with \(\sum_{i \in I} \sum_{t \in T_i} r_{it} y_{it} = 0\), then \(\sum_{t \in T_k} r_{kt} y_{kt} = -\sum_{i \in I \setminus \{k\}} \sum_{t \in T_i} r_{it} y_{it}, \forall k \in I\). But \(M_k \cap (\sum_{i \in I \setminus \{k\}} M_i) = 0\), so \(\sum_{t \in T_k} r_{kt} y_{kt} = 0\). But \((y_{kt})_{t \in T_k}\) is a basis in \(M_k\), so \(r_{kt} = 0, \forall k \in I, \forall t \in T_k\).

4.12 Proposition. Every module is (isomorphic to) a factor module of a free module. More precisely, if \(M\) is an \(R\)-module and \(S\) is a system of generators of \(M\), then there exists an epimorphism \(\varphi: R^S \to M\). Hence \(M \cong R^S / \text{Ker } \varphi\).

Proof. Let \((e_s)_{s \in S}\) be the canonical basis in \(R^S\). Define \(\varphi: R^S \to M\) by \(\varphi(e_s) = s, \forall s \in S\) (see 4.9). Since the submodule \(\text{Im } \varphi\) includes \(S\), and \(S\) generates \(M\), we obtain \(\text{Im } \varphi = M\).

This simple fact is very important: if one knows the structure of the submodules and of the factor modules of free modules, the structure of an arbitrary module is known. This method will be used to study the structure of finitely generated modules over a PID. Here is another application of this principle:

4.13 Example. If \(M\) is an \(R\)-module, then there exists an exact sequence:

\[\ldots \to L_{n+1} \to L_n \to \ldots \to L_1 \to L_0 \to M \to 0,\]

where \(L_n\) is free, \(\forall n \in \mathbb{N}\). Such a sequence is called a free resolution of \(M\). For the proof, let \(L_0\) be free and let \(u_0: L_0 \to M\) be an epimorphism (as given by the Proposition above). Then we have an exact sequence: \(0 \to \text{Ker } u_0 \xrightarrow{i} L_0 \xrightarrow{u_0} M \to 0\). Apply again the proposition for \(\text{Ker } u_0\) and obtain a free module \(L_1\) and an epimorphism \(u_1: L_1 \to \text{Ker } u_0\). We have now the exact sequence.
II. Modules

0 \rightarrow \text{Ker } u_1 \rightarrow L_1 \xrightarrow{i \circ u_1} L_0 \xrightarrow{u_0} M \rightarrow 0

since \text{Ker } u_0 = \text{Im } i \circ u_1. One continues by induction on \( n \) such that exists an exact sequence of the form:

0 \rightarrow K_n \rightarrow L_n \rightarrow ... \rightarrow L_1 \rightarrow L_0 \rightarrow M \rightarrow 0,

with \( L_i \) free, \( 1 \leq i \leq n \).

In a vector space, any two bases have the same cardinal. In the case of a free module over an arbitrary ring, this fact is not guaranteed. Nevertheless, the free modules over a commutative ring have this property; in fact, the proof of this result reduces the problem to the case of vector spaces.

**4.14 Proposition.** Let \( R \) be a commutative ring and let \( M \) be a free \( R \)-module. Then any two bases of \( M \) have the same cardinal.

**Proof.** Let us try to connect the \( R \)-module \( M \) to a vector space over a field. Recall that, if \( I \) is a maximal ideal in \( R \), then \( R/I \) is a field.

A maximal ideal \( I \) of \( R \) exists (see 1.20). Then \( R/I \) is a field (see Appendix, Prime and maximal ideals). The module \( M \) has no natural structure of \( R/I \)-vector space: the natural “multiplication” given by \((r + I)x = rx\), for \( r \in R, x \in M \), is not well defined unless \( ix = 0, \forall i \in I \). This is the reason we need to “kill off” the elements \( x \) of \( M \) for which \( ix \neq 0 \) for some \( i \in I \), as follows:

Let \( IM := \{ a_1x_1 + ... + a_nx_n \mid n \in \mathbb{N}, a_i \in I, x_i \in M, \forall i = 1, ..., n \} \). A straightforward check shows that \( IM \) is an \( R \)-submodule of \( M \) and that the factor module \( M/IM \) is also an \( R/I \)-module (i.e. an \( R/I \)-vector space!) with respect to the external operation:

\((r + I)(x + IM) := rx + IM, \forall r \in R, \forall x \in M. \)

Let \( \pi : M \rightarrow M/IM \) be the canonical surjection, \( \pi(x) = x + IM, \forall x \in M. \) We prove that: if \((e_\alpha)_{\alpha \in A} \) is a basis of \( R \) \( M \), then \((\pi(e_\alpha))_{\alpha \in A} \) is a basis in the \( R/I \)-vector space \( M/IM \).
Indeed, \((\pi(e_\alpha))_{\alpha \in A}\) generates the \(R\)-module \(M/IM\), since \(\pi\) is surjective and \((e_\alpha)_{\alpha \in A}\) generates \(M\). It easily seen that it is also generates \(M/IM\) as an \(R/I\)-module.

Let us show that \((\pi(e_\alpha))_{\alpha \in A}\) is \(R/I\)-linear independent. Let \((r_\alpha + I)_{\alpha \in A}\) be a family of elements of \(R/I\) (where \(r_\alpha \in R, \forall \alpha \in A\)), of finite support, such that

\[
\sum_{\alpha \in A} (r_\alpha + I) \pi(e_\alpha) = 0 + IM.
\]

This means: \(\sum_{\alpha \in A} r_\alpha e_\alpha \in IM\). Note that, for any \(x \in IM, x = a_1x_1 + \ldots + a_mx_m \in IM\), for some \(a_i \in I, x_i \in M, \forall i = 1, \ldots, m\). Writing every \(x_i\) as a linear combination of \((e_\alpha)_{\alpha \in A}\), it follows that \(x\) is of the form \(\sum_{\alpha \in A} b_\alpha e_\alpha\), where \(b_\alpha \in I, \forall \alpha \in A\) (since \(a_i \in I\) and \(I\) is an ideal). Thus,

\[
\sum_{\alpha \in A} r_\alpha e_\alpha = \sum_{\alpha \in A} b_\alpha e_\alpha
\]

But \((e_\alpha)_{\alpha \in A}\) is a basis, so \(r_\alpha = b_\alpha \in I, \forall \alpha \in A\). Thus, \(r_\alpha + I = b_\alpha + I = 0 + I, \forall \alpha \in A\).

Now, taking two bases \((e_\alpha)_{\alpha \in A}\) and \((v_\beta)_{\beta \in B}\) of \(RM\), we obtain that \((\pi(e_\alpha))_{\alpha \in A}\) and \((\pi(v_\beta))_{\beta \in B}\) are bases in the \(R/I\)-vector space \(M/IM\). Since two bases of a vector space have the same cardinal, we obtain \(|A| = |B|\). \(\square\)

**4.15 Definition.** Let \(R\) be a commutative ring and let \(L\) be a free \(R\)-module. The cardinal of a basis of \(RL\) is called the *rank of \(L\)*, denoted \(\text{rank}_R L\) (or, simply, \(\text{rank} L\)). The previous result ensures that the definition is independent of a choice of the basis in \(RL\). If \(R\) is a field, the *rank* of a vector space is the same as its *dimension*.

We study now the *homomorphisms between free modules of finite rank*. The idea is to mimic the vector space situation: for two fixed bases in the modules, we associate a *matrix* to a homomorphism. In this manner, the operations with (and the properties of) module homomorphisms translate in matrix language (and conversely). We...
suppose that the ring $R$ is commutative, although some results hold for any ring with identity.

4.16 Definition. Let $E$ and $F$ be free $R$-modules and let $\varphi : E \to F$ be a homomorphism. Fix $e = (e_1, \ldots, e_m)$ an ordered basis in $E$ and $f = (f_1, \ldots, f_n)$ an ordered basis\textsuperscript{15} in $F$. For any $i \in \{1, \ldots, m\}$, there exist and are unique $a_{ij} \in R$, ($j \in \{1, \ldots, n\}$) such that

$$\varphi(e_i) = a_{i1}f_1 + \ldots + a_{in}f_n.$$ 

The matrix $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in \text{M}_{m,n}(R)$ is denoted by $\text{M}_{e, f}(\varphi)$ and it is called the \textit{matrix of the homomorphism} $\varphi$ \textit{(in the pair of bases} $(e, f)$). 

In other words, the \textit{i-th row} of $\text{M}_{e, f}(\varphi)$ is made up of the coordinates $a_{i1}, \ldots, a_{im}$ of $\varphi(e_i)$ in the basis $f$. If $E = F$, one usually takes $e = f$ and $\text{M}_{e, f}(\varphi)$ is denoted simply $\text{M}_e(\varphi)$. The importance of ordering the bases is quite clear now: a permutation of the basis $(e_1, \ldots, e_m)$ leads to another matrix of $\varphi$, whose rows are a permutation of the rows of the initial matrix.

4.17 Remark. This is the “algebraic” convention of writing the matrix $\text{M}_{e, f}(\varphi)$. Of course, one can agree to the “geometric convention”: write the coordinates of $\varphi(e_i)$ in the basis $f$ on the \textit{i-th column} (which means considering the transpose matrix $^t\text{M}_{e, f}(\varphi)$). Choosing this rule leads to changing property $b)$ in the next sentence into: $^t\text{M}_{e, f}(\psi \circ \varphi) = ^t\text{M}_{f, g}(\psi) \cdot ^t\text{M}_{e, f}(\varphi)$.

4.18 Proposition. \textit{a) The mapping:}

$$\text{M}_{e, f} : \text{Hom}_R(E, F) \to \text{M}_{m,n}(R),$$

$$\varphi \mapsto \text{M}_{e, f}(\varphi)$$

\textit{is an R-module isomorphism.}

\textsuperscript{15} We consider the bases as being \textit{totally ordered} (the place of the element in the basis matters).
II.4 Free modules

b) If \( G \) is a free \( R \)-module, \( g = (g_1, \ldots, g_p) \) is a basis in \( G \), and \( \psi : F \to G \) is an \( R \)-module homomorphism, then \( M_{e, g}(\psi \circ \phi) = M_{e, f}(\phi) \cdot M_{f, g}(\psi) \).

c) The mapping \( M_e : \text{End}_R(E) \to M_m(R) \) is an \( R \)-algebra anti-isomorphism (i.e. \( M_e : \text{End}_R(E)^{\text{op}} \to M_m(R) \) is an \( R \)-algebra isomorphism).

Proof. a) Let \( \eta : E \to F \) be another homomorphism and let \( (b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} = M_{e, \eta} \). Then

\[(\phi + \eta)(e_i) = \phi(e_i) + \eta(e_i) = \sum_j a_{ij} f_j + \sum_j b_{ij} f_j = \sum_j (a_{ij} + b_{ij}) f_j\]

So, \( M_{e, \eta}(\phi + \eta) = (a_{ij} + b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} = M_{e, \eta}(\phi) + M_{e, \eta}(\eta) \), so \( M_{e, \eta} \) is an Abelian group homomorphism. If \( M_{e, \eta}(\phi) = 0 \) (the matrix \( 0 \in M_{m, n}(R) \)), then \( \phi(e_i) = 0, 1 \leq i \leq m \). Since \( (e_i)_{1 \leq i \leq m} \) is a basis in \( E \), \( \phi = 0 \); this shows that \( M_{e, \eta} \) is injective. For the surjectivity, let \( A = (a_{ij}) \in M_{m, n}(R) \). The universality property of the free module yields a unique homomorphism \( \varphi \in \text{Hom}_R(E, F) \) such that \( \varphi(e_i) = a_{in} f_1 + \cdots + a_{in} f_n \), i.e. \( M_{e, \varphi} = A \). If \( r \in R \) and \( \varphi \in \text{Hom}_R(E, F) \), a simple computation shows that \( M_{e, \varphi}(r \varphi) = rM_{e, \varphi}(\varphi) \).

b) Let \( M_{f, g}(\psi) = (b_{jk})_{1 \leq j \leq m, 1 \leq k \leq p} \). We have

\[(\psi \circ \phi)(e_i) = \psi(\sum_j a_{ij} f_j) = \sum_j a_{ij} \psi(f_j) = \sum_j a_{ij} \sum_k b_{jk} g_k = \sum_k (\sum_j a_{ij} b_{jk}) g_k.\]

So, \( M_{e, g}(\psi \circ \phi) = M_{e, \phi}(\psi) \cdot M_{f, g}(\psi) \).

c) The statement follows from a) and b): a) implies that \( M_e \) is an \( R \)-module homomorphism and b) implies that \( M_e : \text{End}_R(E)^{\text{op}} \to M_m(R) \) is a ring homomorphism. \( \square \)

In the setting above, if \( e = (e_1, \ldots, e_n) \) and \( f = (f_1, \ldots, f_n) \) are ordered bases of the free module \( E \), then every element of the basis \( f \) is written uniquely as a linear combination of \( e_i \)'s:

\[f_i = \sum_{j=1}^{n} s_{ij} e_j, s_{ij} \in R\]

We obtain the basis change matrix \( T_{e, f} = (s_{ij}) \in M_n(R) \). If \( g \) is another basis in \( E \), then a straightforward computation yields:

\[T_{e, g} = T_{f, g} \cdot T_{e, f}\]
In particular, \( I_n = T_{e,e} = T_{e,f} \cdot T_{f,e} \), i.e. the basis change matrix is \textit{invertible} in \( M_n(R) \). When is a matrix in \( M_n(R) \) invertible?

If \( A \in M_n(R) \), the \textit{adjoint matrix} of \( A \) is \( A^* \) (the entry \((i,j) \) of \( A^* \) is \((-1)^{i+j} \det(A_{ji}) \), where \( A_{ji} \) is the matrix in \( M_{n-1}(R) \) obtained by suppressing the row \( j \) and the column \( i \) of the matrix \( A \)). Then \( A \cdot A^* = A^* \cdot A = \det(A) \cdot I_n \). This shows that, if \( \det(A) \in U(R) \), then \( A \) is invertible in \( M_n(R) \) and \( A^{-1} = (\det A)^{-1} A^* \). Conversely, if \( A \) is invertible, \( A \cdot A^{-1} = I_n \) implies (taking determinants) that \( \det(A \cdot A^{-1}) = \det(A) \cdot \det(A^{-1}) = 1 \), so \( \det(A) \in U(R) \). Thus, we have:

\textbf{4.19 Proposition.} \textit{Let} \( R \) \textit{be a commutative ring and let} \( n \in \mathbb{N}^* \).

(a) \( A \) \textit{matrix} \( S \in M_n(R) \) is \textit{invertible} in \( M_n(R) \) \textit{if and only if} \( \det(S) \) is \textit{invertible} in \( R \).

(b) \textit{Let} \( E \) \textit{be a free} \( R \)-\textit{module of rank} \( n \) \textit{and let} \( e = \{e_1, \ldots, e_n\} \), \( f, g \) \textit{be bases in} \( E \). \textit{Then}:

\[ T_{e,g} = T_{f,g} \cdot T_{e,f} \]

In particular, any basis change matrix \( T_{e,f} \) is invertible. Conversely, if \( S = (s_{ij}) \) is invertible in \( M_n(R) \) and \( f_i = \sum_{j=1}^{n} s_{ij} e_j \), \( 1 \leq i \leq n \), then \( \{f_1, \ldots, f_n\} \) is a basis in \( E \).

The facts not already proven in the proposition above have proofs that can be taken word for word from the case of the vector spaces.

The multiplicative group of the \( n \times n \) invertible matrices with entries in \( R \), \( U(M_n(R)) = \{S \in M_n(R) \mid \exists T \in M_n(R) \text{ such that } ST = TS = I\} = \{S \in M_n(R) \mid \det S \in U(R)\} \) is also denoted \( GL(n, R) \) or \( GL_n(R) \) and is called the \textit{linear general group of degree} \( n \) \textit{over} \( R \).

The following formalism is useful in many situations involving vectors and matrices. Let \( R \) be a commutative ring and let \( E \) be an \( R \)-module. If \( m, n, p \) are positive integers, let \( M_{n,p}(E) \) denote the set of \( n \times p \) matrices with entries in \( E \). It is evident that \( M_{n,p}(E) \) is an Abelian
group with respect to usual matrix addition. For any 
\( A = (a_{ij}) \in M_{m,n}(R) \) and any 
\( X = (x_{jk}) \in M_{n,p}(E) \), define the product 
\( AX \in M_{m,p}(E) \) by
\[
AX = (y_{ik}) \in M_{m,p}(E), \quad y_{ik} = \sum_{j=1}^{n} a_{ij}x_{jk}, \quad \forall i \in \{1, \ldots, m\}, \forall k \in \{1, \ldots, p\},
\]

Of course, \( a_{ij}x_{jk} \) is the product given by the 
\( R \)-module structure of \( E \).

We obtain an “external operation”
\[
\cdot : M_{m,n}(R) \times M_{n,p}(E) \to M_{m,p}(E),
\]
with the following properties (the proof of which is straightforward):
\[
(A + B)X = AX + BX, \quad A(X + Y) = AX + AY, \quad \forall A, B \in M_{m,n}(R), \forall X, Y \in M_{n,p}(E).
\]

Moreover, if \( q \in \mathbb{N}^* \), then for any 
\( A \in M_{q,m}(R), B \in M_{m,n}(R), X \in M_{n,p}(E) \) we have:
\[
(AB)X = A(BX).
\]

If \( I \) is the identity matrix in \( M_{n}(R) \), then 
\( IX = X, \forall X \in M_{n,p}(E) \).

Thus, \( M_{n,p}(E) \) is a left module over the ring \( M_{n}(R) \).

As an application, let us see how the matrix of a homomorphism of 
free modules changes when we change bases in the modules.

4.20 Proposition. Let \( E \) and \( F \) be free \( R \)-modules, \( \text{rank } E = n, \text{rank } F = m \) and let 
\( e = (e_1, \ldots, e_n), e' = (e'_1, \ldots, e'_n) \) be bases in \( E \),
\( f = (f_1, \ldots, f_m), f' = (f'_1, \ldots, f'_m) \) bases in \( F \). Let \( S = (s_{ij}) \in M_{n}(R) \) be the 
basis change matrix from \( e \) to \( e' \) and let \( T = (t_{ij}) \in M_{m}(R) \) be the basis 
change matrix from \( f \) to \( f' \). If \( \phi : E \to F \) is a homomorphism, then:
\[
M_{e,f}(\phi) = S \cdot M_{e',f'} \cdot T^{-1}.
\]

Proof. Let \( M_{e,f}(\phi) = A = (a_{ij}) \in M_{n,m}(R) \), i.e.:
\[
\phi(e_i) = \sum_{j} a_{ij} f_j, \quad \forall i \in \{1, \ldots, n\}.
\]

We view \( e = ^{t}(e_1, \ldots, e_n) \) as a column matrix, \( e \in M_{n,1}(E) \) and, similarly, \( f \in M_{m,1}(F) \). The relations above can be written as
\[
\phi(e) = A \cdot f,
\]
where $\varphi(e) = (\varphi(e_1), \ldots, \varphi(e_n))$. In the same way, if $B = M_{e'}f(\varphi)$, then $\varphi(e') = B \cdot f'$. Also we have $e' = S \cdot e$ and $f' = T \cdot f$, or $f = T^{-1} \cdot f'$. Since $\varphi$ is an $R$-module homomorphism, we have $\varphi(S \cdot e) = S \cdot \varphi(e)$ (prove this!).

Thus:

$$\varphi(e') = \varphi(S \cdot e) = S \cdot \varphi(e) = S \cdot (A \cdot f) = (SA) \cdot f = (SA) \cdot (T^{-1} \cdot f') = (SAT^{-1}) \cdot f',$$

which says that the matrix of $\varphi$ in the bases $e'$ and $f'$ is $SAT^{-1}$.  

\begin{center}
\textbf{Exercises}
\end{center}

1. Let $R$ be a ring and let $L$ be a free $R$-module. Then any $R$-epimorphism $\varphi : M \to L$ splits. In particular, if $K$ is a field, any short exact sequence of $K$-linear spaces $0 \to U \to V \to W \to 0$, is split.

2. a) Let $R$ be a domain and $g_1, \ldots, g_n \in R[X]$, with $\deg g_i \neq \deg g_j$ if $i \neq j$. Then $g_1, \ldots, g_n$ are linearly independent in the $R$-module $R[X]$.

b) Let $K$ be a field of characteristic 0 (i.e. $n \cdot 1 \neq 0$, $\forall n \in \mathbb{N}^*$, where 1 is the identity of $K$) and let $a \in K$. Then $\{1, X - a, (X - a)^2, \ldots, (X - a)^n, \ldots\}$ is a basis in the $K$-vector space $K[X]$. If $p \in K[X]$, compute the coordinates of $p$ in this basis. (\textit{Hint}. Recall the Taylor series expansion, applied to polynomials.)

3. Prove that in the $\mathbb{Z}$-module $\mathbb{Q}$ any subset having at least two elements is linearly dependent and that $\mathbb{Z} \not\cong \mathbb{Z} \mathbb{Q}$. Deduce that $\mathbb{Z} \mathbb{Q}$ is not free.

4. Let $G$ be a finite Abelian group. Can $G$ be a free $\mathbb{Z}$-module?

5. Characterize the ideals $I$ of the commutative ring $R$ that are free $R$-modules.
6. If $M$ is a free $R$-module of basis $B$, express $|M|$ as a function of $|R|$ and $|B|$.

7. Let $M$ be an $R$-module with the following property:

There exists a subset $B$ of $M$ such that, for any $_R N$ and any function $\varphi : B \to N$, there exists a unique $R$-homomorphism $\psi : M \to N$ with $\psi|_B = \varphi$.

Then $M$ is free and $B$ is a basis of $M$. (the converse of Prop. 4.9).

8. Let $R$ be a commutative ring and let $I, J$ be ideals in $R$. Consider the statements:

(i) $I \cong J$ (as $R$-modules).
(ii) $R/I \cong R/J$ (as $R$-modules).
(iii) $R/I \cong R/J$ (as rings).

Which are the valid implications between the above statements?

9. Let $W$ be a finite dimensional $K$-vector space and let $U, V \leq kW$.

a) Show that $\dim(W/U) = \dim W - \dim U$.

b) Using the isomorphism theorems, prove that $\dim(U + V) = \dim U + \dim V - \dim(U \cap V)$.

10. Is it true that any submodule (resp. factor module) of a free module is still a free module?
III. Finitely generated modules over principal ideal domains

In the case of finitely generated modules over PID's very precise structure theorems are available. Applying these theorems for $\mathbb{Z}$, a complete description of finitely generated Abelian groups is obtained. Another important application is to the problem of invariant subspaces of an endomorphism of a finitely dimensional vector space: we obtain the existence of a basis in which the given endomorphism has a “simple” matrix (the Jordan canonical form).

III.1 The submodules of a free module

Any module is a factor module of a free module. In order to build a factor module we need a submodule. It is therefore natural to study first the submodules of a free module. Recall that, if $R$ is commutative, any free $R$-module $E$ has the property that any two bases have the same cardinal (called the rank of $E$).

In this section, unless otherwise specified, $R$ denotes a principal ideal domain.
1.1 Theorem. Let $E$ be a free $R$-module of rank $n$ and let $F$ be a submodule of $E$. Then $F$ is free and rank $F \leq n$.

Proof. Evidently, we may suppose $F \neq 0$.

We use an argument by induction on $n$. If $n = 1$ let \{e\} be a basis of $E$. Then $E = Re \cong R$. In this case, the statement of the theorem becomes: any submodule (= ideal) of $R$ is free, of rank $\leq 1$. Since $R$ is a PID, this is true: any nonzero ideal of $R$ is of the form $Ra$, with $a \in R$; so \{a\} is a basis of $Ra$.

Suppose that, for any free $R$-module $H$ of rank $n - 1$, any submodule of $H$ is free, of rank $\leq n - 1$. Take $R E$, free of rank $n$, \{e$_1$, \ldots, e$_n$\} a basis in $E$ and $F \leq R E$. Let $L := Re_2 + \ldots + Re_n$ and $G := F \cap L$. Then $L$ is free, of basis \{e$_2$, \ldots, e$_n$\}, so its rank is $n - 1$, and $G = F \cap L \leq L$. By induction, $G$ is free of rank $m \leq n - 1$. If $G = F$, we are done. If not, note that $F + L \leq R E$, so

\[
\frac{F}{G} \cong \frac{F}{L \cap F} \cong \frac{F + L}{L} \leq \frac{E}{L}
\]

Accordingly, $0 \neq F/G$ is isomorphic to a submodule of $E/L$. But $E/L$ is free, of basis \{e$_1 + L$\} (easy check) and thus $F/G$ is free of rank $1$ (by the case $n = 1$, already proven). Let \{f$_1$, \ldots, f$_m$\} be a basis in $G$ and \{f$_0 + G$\} a basis in $F/G$. We claim that $B = \{f$_0$, f$_1$, \ldots, f$_m$\}$ is a basis in $F$.

Indeed, if $a_0$, a$_1$, \ldots, a$_m \in R$, with $a_0 f_0 + a_1 f_1 + \ldots + a_m f_m = 0$, then $a_0 f_0 + G = 0 + G$ (since $a_1 f_1 + \ldots + a_m f_m \in G$); but \{f$_0 + G$\} is a basis in $F/G$, so $a_0 = 0$. We get $a_1 f_1 + \ldots + a_m f_m = 0$, which implies $f_1 = \ldots = f_m = 0$ (\{f$_1$, \ldots, f$_m$\} being a basis in $G$). Therefore, $B$ is linearly independent.

Let us show that it is a generating system. Let $x \in F$. Then $x + G \in F/G$, so there is some $a_0 \in R$ such that $x + G = a_0 f_0 + G$. This means that $x - a_0 f_0 \in G$, so $x - a_0 f_0 = a_1 f_1 + \ldots + a_m f_m$, for some $a_i \in R$. Thus, $x = a_0 f_0 + a_1 f_1 + \ldots + a_m f_m$. \qed
Let $R E$ and $F \leq R E$, as above. If $\{e_1, \ldots, e_n\}$ is a basis of $E$ and $\{f_1, \ldots, f_m\}$ is a basis of $F$, any $f_i$ is written uniquely as a linear combination of $\{e_1, \ldots, e_n\}$:

$$f_i = \sum_j a_{ij} e_j, \text{ with } a_{ij} \in R.$$  

We obtain a matrix $A = (a_{ij}) \in M_{m, n}(R)$. Can we choose the bases in $E$ and $F$ such that $A$ has a simple (e.g., diagonal) form?\(^1\) The next theorem says the answer is Yes.

**1.2 Theorem.** Let $R$ be a principal ideal domain, $E$ a free $R$-module of rank $n$ and $F$ a submodule of $E$. Then there exists a basis $e = (e_1, \ldots, e_n)$ in $E$ and a basis $f = (f_1, \ldots, f_m)$ in $F$, such that $m \leq n$, $f_i = d_i e_i, \forall i \in \{1, \ldots, m\}$, where $d_1, \ldots, d_m \in R$ and $d_1 | d_2 | \cdots | d_m$.

**(Beginning of) Proof.** For any basis $e' = (e'_1, \ldots, e'_n)$ in $E$ and any basis $f' = (f'_1, \ldots, f'_m)$ in $F$, there exists a unique matrix $A = (a_{ij}) \in M_{m, n}(R)$ such that $f'_i = \sum_j a_{ij} e'_j, \forall i \in \{1, \ldots, m\}$. If we see $e'$ as a column matrix, $e' \in M_{n, 1}(E)$, (see the discussion preceding 4.20) and $f'$ as a matrix in $M_{m, 1}(R)$, then these relation are written shortly:

$$f' = Ae'.$$

Suppose the problem is solved: we have two bases $e = (e_1, \ldots, e_n)$ and $f = (f_1, \ldots, f_m)$ as in the statement of the theorem. Then the unique matrix $D = (d_{ij}) \in M_{m, n}(R)$ with $f = De$ has the properties: $d_{ij} = 0$ if $i \neq j$ and $d_{11} | d_{22} | \cdots | d_{mm}$. Let $V \in \text{GL}_n(R)$ the basis change matrix from basis $e$ to basis $e'$ and $U \in \text{GL}_m(R)$ the basis change matrix from basis $f$ to basis $f'$. Thus, $e' = Ve$ and $f' = Uf$. Replacing in $f' = Ae'$, we have $Uf = A(Ve) = (AV)e$; so

$$f = U^{-1}(AV)e.$$  

Comparing with $f = De$, we obtain $D = U^{-1}AV$.

We see that the existence of the bases $e$ and $f$ with the desired property is equivalent to the following: given $A \in M_{m, n}(R)$, there exist

\(^1\) What about the case of vector spaces?
invertible matrices $U$ and $V$ such that $U^{-1}AV = D = (d_{ij}) \in \text{M}_{m,n}(R)$ is a diagonal matrix ($i \neq j$ implies $d_{ij} = 0$), with the additional condition $d_{11}|d_{22}| \ldots |d_{mm}$.

We reduced the statement to a problem concerning matrices with elements in $R$. The following definitions are helpful:

1.3 Definitions. Let $m, n \in \mathbb{N}^*$.

a) A matrix $D = (d_{ij}) \in \text{M}_{m,n}(R)$ is called a diagonal matrix if its entries not situated on the main diagonal are zero: $i \neq j$ implies $d_{ij} = 0$. If $r = \min(m,n)$ and $d_1, \ldots, d_r \in R$,

$$\text{diag}(d_1, \ldots, d_r) := \begin{bmatrix} d_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & d_r & 0 \end{bmatrix}$$

denotes the diagonal matrix $(d_{ij}) \in \text{M}_{m,n}(R)$ with $d_{ii} = d_i$, $\forall i \in \{1, \ldots, r\}$.

b) We call a diagonal matrix $D = \text{diag}(d_1, \ldots, d_r) \in \text{M}_{m,n}(R)$ (where $r = \min(m,n)$) canonically diagonal (or in Smith normal form) if $d_1|d_2| \ldots |d_r$.

c) The matrices $A, B \in \text{M}_{m,n}(R)$ are called arithmetically equivalent (denoted $A \sim B$) if there exist invertible matrices $U \in \text{M}_m(R)$ and $V \in \text{M}_n(R)$ such that $B = UAV$. This is an equivalence relation on $\text{M}_{m,n}(R)$ (exercise!).

We finish the proof of theorem 1.2 by proving the next theorem, which is interesting on its own: the proof is in fact an algorithm used to compute the Smith normal form of a matrix.

1.4 Theorem. Let $R$ be a PID and let $m, n \in \mathbb{N}^*$. Then any matrix $A \in \text{M}_{m,n}(R)$ is arithmetically equivalent to a matrix in Smith normal form. Moreover, the Smith normal form of $A$ is unique in the following sense: if $D = \text{diag}(d_1, \ldots, d_r)$ and $D' = \text{diag}(d'_1, \ldots, d'_r)$ are in Smith nor-
mal form and are arithmetically equivalent, then \( d_1 \sim d_1', \ldots, d_r \sim d_r' \) ("\( \sim \)" means here "associated in divisibility").

If \( D \) is in Smith normal form and is arithmetically equivalent to \( A \), \( D \) is called the Smith normal form of \( A \); it is uniquely determined up to an association in divisibility of the entries on the diagonal.

Before proving the theorem, let us review the transformations that can be performed on the matrix \( A \), such that the resulting matrix is arithmetically equivalent to \( A \). We shall see that these are the transformations that arise when computing determinants: swapping rows (or columns), addition to a row (column) of another row (column) multiplied by some element.

1.5 Definitions. Let \( m \in \mathbb{N}^* \) and let \( I \) be the identity matrix of the ring \( M_m(R) \). An elementary matrix is a square matrix in \( M_m(R) \) which is of one of the following types:

- **Type I**: \( T_{ij}(a) \), where \( a \in R, i, j \in \{1, \ldots, m\}, i \neq j \). \( T_{ij}(a) \) is obtained from \( I \) by adding to the row \( i \) the row \( j \) multiplied by \( a \).

\[
T_{ij}(a) = \begin{bmatrix}
1 & \vdots & \vdots & 0 \\
\vdots & \ddots & \vdots & \vdots \\
i & \cdots & 1 & \cdots & a \\
\vdots & \ddots & \vdots & \vdots \\
j & \cdots & 0 & \cdots & 1 \\
0 & \cdots & \cdots & 1
\end{bmatrix}
\]

- **Type II**: \( P_{ij} \), where \( i, j \in \{1, \ldots, m\}, i \neq j \). \( P_{ij} \) is the matrix obtained from \( I \) by swapping row \( i \) with row \( j \).

- **Type III**: \( D_i(u) \), where \( i \in \{1, \ldots, m\} \) and \( u \in U(R) \). \( D_i(u) \) is the matrix obtained from \( I \) by multiplying row \( i \) by \( u \).
If $A \in M_{m,n}(R)$, and $T_{ij}(a)$, $P_{ij}$, $D_i(u)$ are elementary matrices in $M_{m}(R)$ as above, a direct calculation shows that:
- $T_{ij}(a)A$ is obtained from $A$ by adding to the row $i$ the row $j$ multiplied by $a$.
- $P_{ij}A$ is obtained from $A$ by swapping the row $i$ with the row $j$.
- $D_i(u)A$ is obtained from $A$ by multiplying the row $i$ by $u$.

The transformations described above are called elementary transformations of the rows of $A$ (of type I, II, respectively III). If we take elementary matrices in $M_{n}(R)$ and multiply $A$ on the right with these matrices, we obtain the elementary transformations of the columns of $A$:
- Type I: $AT_{ij}(a)$ is obtained from $A$ by adding to the column $i$ the column $j$ multiplied by $a$.
- Type II: $AP_{ij}$ is obtained from $A$ by swapping the column $i$ with the column $j$.
- Type III: $AD_i(u)$ is obtained from $A$ by multiplying the column $i$ by $u$.

All elementary matrices are invertible. This follows from the following relations, easy to prove:

$$T_{ij}(a)T_{ij}(b) = T_{ij}(a + b), \forall a, b \in R; \text{ so } T_{ij}(a)T_{ij}(-a) = T_{ij}(0) = I.$$
\[ P_{ij}P_{ij} = I; \]
\[ D_{ij}(u)D_{ij}(v) = D_{ij}(uv), \; \forall u, v \in U(R); \text{ so } D_{ij}(u)D_{ij}(u^{-1}) = D_{ij}(1) = I. \]

In other words, the inverse of an elementary matrix exists and is also an elementary matrix (of the same type).

Looking at the definition of the relation of arithmetic equivalence between matrices, we obtain:

**1.6 Proposition.** For any \( A \in M_{m,n}(R) \), any matrix obtained from \( A \) by elementary transformations of rows and/or columns is arithmetically equivalent to \( A \).

We also need:

**1.7 Definition.**
a) Let \( R \) be an UFD and \( a \in R, \; a \neq 0 \). Define the number \( l(a) \), called the length of \( a \), as follows: if \( a \) is a unit, set \( l(a) = 0 \); if \( a \) is nonzero and not a unit, \( l(a) \) is the number of prime factors (not necessarily distinct) of a decomposition of \( a \) in prime factors. For example, in \( \mathbb{Z} \), \( l(1) = 0; \; l(8) = 3; \; l(24) = 4 \). By convention, \( l(0) = -\infty \).

b) If \( A = (a_{ij}) \in M_{m,n}(R) \), define the length of \( A \):

\[ l(A) := \min\{l(a_{ij}) \mid i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}\}. \]

We can finally **prove theorem 1.4:**

The proof is by induction on \( m \). More precisely, we prove that \( P(m) \) holds for any \( m \), where:

**\( P(m) \): For any matrix \( A \in M_{m,n}(R) \), there exists \( d_1 \in R \) and a matrix \( A' \in M_{m-1,n-1}(R) \) such that \( d_1 \) divides all the entries of \( A' \) and \( A \) is arithmetically equivalent to the matrix (written in block form):

\[
\begin{bmatrix}
    d_1 & 0_{1,n-1} \\
    0_{m-1,1} & A'_{m-1,n-1}
\end{bmatrix},
\]

where the subscripts indicate the dimensions of the matrix.

We prove \( P(m) \) by induction on \( l(A) \).
If $A = 0 \ (l(A) = -\infty)$, $A$ is in Smith normal form.

**Case 1.** $A$ has an entry that is a unit ($\iff l(A) = 0$).

Let $a_{ij}$ be a unit. Swap the rows $i$ and 1, then swap the columns $j$ and 1. The matrix obtained is arithmetically equivalent to $A$, with $a_{ij}$ in the position $(1,1)$. To simplify notations, we may suppose thus from the beginning that $a_{11}$ is invertible. We now obtain a matrix that has 0 on the column 1 (except for $a_{11}$): for each $i$, $2 \leq i \leq m$, add to the row $i$ the row 1 multiplied by $(-a_{11}^{-1}a_{i1})$ (which amounts to multiply $A$ on the left with $T_{i1}(-a_{11}^{-1}a_{i1})$). Similarly, we make 0's on the first row and obtain a matrix of the form:

$$B = \begin{bmatrix}
  a_{11} & 0 \\
  0 & A'
\end{bmatrix}$$

with $B \sim A$. The condition that $a_{11}$ divides all entries of $A'$ is fulfilled ($a_{11}$ is invertible!).

**Case 2.** $l(A) \geq 1$.

Let $a_{ij}$ be the entry of $A$ for which $l(a_{ij}) = l(A)$. As in case 1, we may suppose (swapping some rows and/or some columns) that $l(a_{11}) = l(A)$.

**Subcase 2.1.** $a_{11}$ divides all the entries of $A$.

This case is essentially Case 1. Indeed, since $a_{11} | a_{ij}, \exists b_i \in R$ with $a_{ij} = a_{11}b_i$. For $2 \leq i \leq m$, add to the row $i$ the row 1 multiplied by $(-b_i)$; the matrix obtained has 0 on the first column, except $a_{11}$. Clearly, $a_{11}$ divides the entries of the new matrix (these entries are linear combinations of the entries of $A$). In the same way we can annihilate the entries of the first row. Thus, $A$ is arithmetically equivalent to a matrix of the form $B$, as in case 1.

**Subcase 2.2.** $a_{11}$ does not divide all the entries of $A$.

In this case we prove that $A$ is arithmetically equivalent to a matrix $C \in M_{m,n}(R)$ with $l(C) < l(A)$, which finishes the proof of $P(m)$ by induction on $l(A)$.

We may suppose that $a_{11}$ does not divide an entry of the first row or of the first column. Indeed, otherwise $a_{11}$ divides all entries of the first
III. Finitely generated modules over principal ideal domains

126

row and of the first column. Working as in subcase 2.1, we obtain a matrix with $a_{11}$ the only nonzero entry of the first row and of the first column. If $a_{11}$ divides all the entries of the matrix, we are done! If $a_{11}$ does not divide all the entries of the matrix, there exists an entry $a_{ij}$ (with $i, j > 1$) such that $a_{11} \nmid a_{ij}$. By adding to the column 1 the column $j$, we obtain an entry on the column 1, not divisible by $a_{11}$.

To fix the ideas, suppose that $a_{11}$ does not divide an element on the first column (the proof in the other case is similar, the difference being that one multiplies on the right with adequate invertible matrices – see below). Suppose $a_{11} | a_{21}$ (if not, swap rows to achieve this). Let $d = \gcd(a_{11}, a_{21})$. We cannot have $d \sim a_{11}$, because this implies $a_{11} | a_{11}$, false. Since $d | a_{11}$, $l(d) < l(a_{11}) = l(A)$. We exhibit a matrix $C$, arithmetically equivalent to $A$, whose $(1, 1)$-entry is $d$ (thus $l(C) \leq l(d) < l(A)$). Let $a_{11} = da$, $a_{21} = db$, for some $a, b \in R$. Since $(a, b) = 1$ and $R$ is a PID, there exist $u, v \in R$ such that $au + bv = 1$. So, $dau + dbv = a_{11}u + a_{21}v = d$. Consider the matrix (written in block form):

$$
U = \begin{bmatrix}
    u & v & 0 \\
    -b & a & 0 \\
    0 & 0 & I
\end{bmatrix}
$$

where $I$ is the $(m - 2) \times (m - 2)$ identity matrix (if $m = 2$, then $I$ does not appear anymore, i.e. $U$ is a $2 \times 2$ matrix). The matrix $U$ is invertible ($\det U = ua + vb = 1$), so $C := UA \sim A$. But $C$ has as the $(1, 1)$-entry $ua_{11} + va_{21} = d$, so $l(C) \leq l(d) < l(a_{11}) = l(A)$.

This finishes the proof of the existence part.

We prove now the uniqueness part in theorem 1.4. For any $A \in \text{M}_{m,n}(R)$ and $1 \leq k \leq \min(m, n)$, let $\Delta_k(A)$ be the GCD of the minors of order $k$ of the matrix $A$ (Recall that a minor of order $k$ of $A$ is the determinant of a matrix obtained as follows: choose $k$ rows and $k$ columns of $A$ and retain only the entries located at the intersection of the
III.1 The submodules of a free module

chosen rows and columns. There are \( \binom{m}{n} \binom{n}{k} \) minors of order \( k \) in an \( m \times n \) matrix.

Note that, if \( U \in \mathbb{M}_m(\mathbb{R}) \), then \( \Delta_k(A) | \Delta_k(UA) \). Indeed, the rows of \( UA \) are linear combinations (with coefficients in \( \mathbb{R} \)) of the rows of \( A \). Thus, the rows of a minor of order \( k \) of \( UA \) (corresponding to the choice of columns \( i_1, \ldots, i_k \) of \( UA \)) are linear combinations of the rows of \( A \) (truncated to contain only the entries on the columns \( i_1, \ldots, i_k \)). Applying the fact that the determinant of a matrix is a multilinear function of the rows of the matrix\(^2\), it follows that a minor of order \( k \) of \( UA \) is a linear combination of minors of order \( k \) of \( A \). The claim now follows.

Similarly, if \( V \in \mathbb{M}_n(\mathbb{R}) \), then \( \Delta_k(A) | \Delta_k(AV) \). So, if \( A \sim B \), then \( \Delta_k(A) | \Delta_k(B) \) and, by symmetry, \( \Delta_k(B) | \Delta_k(A) \), i.e. \( \Delta_k(A) \sim \Delta_k(B) \). If \( D = \text{diag}(d_1, \ldots, d_r) \) is in Smith normal form, an easy check show that \( \Delta_k(D) \sim d_1 \ldots d_k \). So, if \( A \sim D \), \( d_1, \ldots, d_r \) are determined (up to association in divisibility) by \( \Delta_1(A), \ldots, \Delta_r(A) \) and

\[
d_1 \sim \Delta_1(D) \sim \Delta_1(A), \quad d_k \sim \Delta_k(A) / \Delta_{k-1}(A), \quad \text{for } k \geq 2.
\]

These relations indicate another method to compute effectively \( d_1, \ldots, d_r \) (although the amount of computation is prohibitive if \( m, n \) are not small).

1.8 Remark. The existence part of the proof above is in fact an algorithm to find the Smith normal form of a given matrix. In practice, \( R \) is an Euclidian ring. In this case, the function \( \text{length} \) defined at 1.7 (whose computation is costly, involving the prime factorizations of the

---

\(^2\) By denoting \( (l_1, \ldots, l_k) \) the matrix having the rows \( l_1, \ldots, l_k \), the following relation holds: \( \det(al_1 + bl_1', \ldots, l_k) = a\det(l_1, \ldots, l_k) + b\det(l_1', \ldots, l_k) \), \( \forall a, b \in \mathbb{R} \) (similarly for any row \( l_i \)).
elements of the matrix) can be advantageously replaced by the function $\varphi$ in the definition of the Euclidian ring.

To be precise, if $R$ is Euclidian with respect to $\varphi$, and $A = (a_{ij}) \in M_{m,n}(R)$, define

$$\varphi(A) = \min\{ \varphi(a_{ij}) \mid i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\} \}.$$ 

The proof above is rewritten word for word (replace everywhere $l$ with $\varphi$); at the subcase 2.2, replace the matrix $U$ with $T_{21}(−q)$, where $a_{21} = a_{11}q + r$, with $\varphi(r) < \varphi(a_{11}) = \min\{ \varphi(a_{ij}) \}$ (i.e., subtract from the row 2 of $A$ the row 1 multiplied by $q$, the quotient in the division with remainder of $a_{21}$ to $a_{11}$). This places $r$ on the entry $(2, 1)$ of the matrix $UA$, which has thus $\varphi(UA) < \varphi(A)$. We invite the reader to check the details and apply the algorithm in the proof in concrete cases (see also the exercises).

### Exercises

In the exercises, $R$ denotes a principal ideal domain, unless otherwise specified.

**1.** Find the Smith normal form of the matrix:

$$\begin{bmatrix} 2 & 6 & 9 \\ 5 & 10 & 12 \\ 0 & 6 & 12 \end{bmatrix} \in M_3(\mathbb{Z}).$$

If $L$ is the submodule of $\mathbb{Z}^3$ generated by $v_1 = (2, 6, 9)$, $v_2 = (5, 10, 12)$, $v_3 = (0, 6, 12)$, determine a basis of $L$, rank $L$ and the factor module $\mathbb{Z}^3/L$. 
2. Let \( a, b \in R \). Show that the Smith normal form of \[
\begin{bmatrix}
a & 0 \\
0 & b
\end{bmatrix}
\] is \[
\begin{bmatrix}
d & 0 \\
0 & m
\end{bmatrix},
\] where \( d = \text{GCD}(a, b) \), \( m = \text{LCM}(a, b) \). (Hint. Use the invariants \( \Delta_k \).)

3. Find the Smith normal form of a diagonal matrix \( \text{diag}(a_1, \ldots, a_n) \in M_n(R) \).

4. Find the Smith normal form of a row matrix \( (a_1, \ldots, a_n) \in R) \).

5. Determine all subgroups of \((\mathbb{Z} \times \mathbb{Z}, +)\).

6. Let \( n \in \mathbb{N}^* \), \( x_1, \ldots, x_n \in R \) and \( d = \text{GCD}(x_1, \ldots, x_n) \). Show that there exists \( V \in \text{GL}(n, R) \) such that \( (x_1, \ldots, x_n)V = (d, 0, \ldots, 0) \). (Ind. Consider the Smith normal form of the row matrix \((x_1, \ldots, x_n))\).

7. Let \( n \in \mathbb{N}^* \) and let \( a_1, \ldots, a_n \in R \). Show that: there exists \( V \in \text{GL}(n, R) \) such that the first row of \( V \) is \((a_1, \ldots, a_n)\) if and only if \( \text{GCD}(a_1, \ldots, a_n) = 1 \).

8. Let \( K \) be a field and let \( A \in M_{m,n}(K) \). Then the Smith normal form of \( A \) is \( \text{diag}(1, \ldots, 1, 0, \ldots, 0) \), where 1 appears \( r \) times (\( r \) is the rank of the matrix \( A \)).

9. Let \( m, n \in \mathbb{N}^* \) and let \( \varphi : R^n \rightarrow R^m \) be an \( R \)-homomorphism whose matrix is \( A \in M_{m,n}(R) \) (in the canonical bases). Let \( U \in \text{GL}(m, R) \) and \( V \in \text{GL}(n, R) \) such that \( UAV \) is in Smith normal form, namely \( \text{diag}(d_1, \ldots, d_r, 0, \ldots, 0) \), with \( r \leq \min(m, n) \) and \( d_1, \ldots, d_r \) nonzero. Show that a basis in \( \text{Ker}\varphi \) is \((v_{r+1}, \ldots, v_n)\), where \( v_i \) is the column \( i \) of the matrix \( V \) (\( v_i \) is seen as an element in \( R^n \)).

10. Suppose \( L \) is a free \( R \)-module of rank \( n \), \((e_1, \ldots, e_n)\) is a basis in \( L \) and \( \{f_1, \ldots, f_m\} \subseteq L \). Show that a basis in \( F = \langle f_1, \ldots, f_m \rangle \) can be obtained as follows:
Let \( A = (a_{ij}) \in M_{m, n}(R) \) such that \( f_i = \sum_j a_{ij} e_j \) (\( 1 \leq i \leq n \)) and let \( U \in GL_m(R), \ V \in GL_n(R), \) with \( UAV \) in Smith normal form. Let \( g_i := \sum_j u_{ij} f_j \in F. \) Then a basis in \( F \) is \( \{g_i \mid 1 \leq i \leq m, \ g_i \neq 0\}. \)

11. Let \( m, n \in \mathbb{N}^* \), \( A \in M_{m, n}(R) \) and \( b \in M_{m, 1}(R) \). Consider the equation

\[
Ax = b, \ x \in R^n
\]

(a linear system of \( m \) equations and \( n \) unknowns). Consider the “extended matrix” \( \bar{A} = (A, b) \in M_{m, n+1}(R) \) (the first \( n \) columns of \( \bar{A} \) are the columns of \( A \), the last column is \( b \)). Show that: \( Ax = b \) has solutions \( x \) in \( R^n \) if and only if the Smith normal form of \( \bar{A} \) is \( (D, 0) \), where \( D \) is the Smith normal form of \( A \) and 0 is the zero column in \( M_{m, 1}(R) \). Note that if \( R \) is a field, this is the Kronecker-Capelli theorem. (see also exercise 8.)

### III.2 Finitely generated modules over a principal ideal domain

We are now ready to describe the structure of finitely generated modules over a principal ideal domain. Since any module is a factor of a free module, and a detailed description of any submodule of a finitely generated free module is available (Theorem 1.2), the task is now easy. We need only some preparations:

2.1 Definition. If \( R \) is a PID, \( M \) is an \( R \)-module and \( x \in M \), then \( \text{Ann}_R(x) = \{r \in R \mid rx = 0\} \) is a principal ideal of \( R \). A generator of the ideal \( \text{Ann}_R(x) \) is called an order of \( x \), denoted by \( o(x) \). Therefore, the order of an element in \( M \) is defined up to an association in divisibility. Thus, we have
Ann_R(x) = Ro(x); Rx \cong R/Ro(x).

This notion generalizes the usual concept of order of an element in an Abelian group.

If \( \phi : E \rightarrow F \) is an \( R \)-module isomorphism and \( x \in E \), then
\[ o(x) = o(\phi(x)) \]
since \( \text{Ann}_R(x) = \text{Ann}_R(\phi(x)) \).

### 2.2 Lemma

Let \( R \) be a ring and let \( M \) be a left \( R \)-module such that
\( M \) is the direct sum of a family of submodules \( (M_i)_{i \in I} \), \( M = \bigoplus M_i \). If \( N_i \leq_R M_i \), \( \forall i \in I \), then the sum of the submodules \( (N_i)_{i \in I} \) is direct and we have a canonical isomorphism:
\[ \bigoplus_{i \in I} M_i \cong \bigoplus_{i \in I} N_i. \]

**Proof.** Let \( \pi_j : \bigoplus_I M_i \rightarrow M_j \) be the canonical surjections. Define \( \phi : M \rightarrow \bigoplus_{i \in I} M_i \) by \( \phi(x) = (\pi_i(x) + N_i)_{i \in I}, \ \forall x \in M \). One easily checks that \( \phi \) is a surjective homomorphism (in fact, \( \phi \) is the direct sum of the family of homomorphisms \( \eta_i \circ \pi_i : M \rightarrow M_i/N_i \), where \( \eta_i : M_i \rightarrow M_i/N_i \) is the canonical surjection). We have \( \text{Ker}\phi = \{ x \in M | \pi_i(x) \in N_i, \ \forall i \in I \} \). Since \( x = \sum_{i \in I} \pi_i(x) \), it follows that \( \text{Ker}\phi = \sum_{i \in I} N_i = \bigoplus_I M_i \).
Apply now the isomorphism theorem. \( \square \)

We state now the following important theorem, which determines the structure of finitely generated modules over a principal ideal domain. Recall that \( R^\circ \) is the set of nonzero noninvertible elements of \( R \).

### 2.3 Theorem

(Invariant factors theorem) Let \( R \) be a principal ideal domain and let \( M \) be a finitely generated \( R \)-module. Then \( M \) is a direct sum of a finite number of cyclic submodules.

More precisely, there exist \( n, m \in \mathbb{N} \), with \( m \leq n \), and \( x_1, ..., x_n \in M \) such that:
\[ M = Rx_1 \oplus ... \oplus Rx_m \oplus Rx_{m+1} ... \oplus Rx_n, \quad (D) \]
Moreover, \( o(x_i) =: d_i \in R \) satisfy the conditions:
\[ d_i \in R^\circ, \forall i \in \{1, \ldots, m\} \text{ and } d_1|d_2|\ldots|d_m; d_{m+1} = \ldots = d_n = 0. \]

The numbers \( n, m \in \mathbb{N} \) and the orders \( o(x_i) \in R, i = 1, n \) with the above properties are uniquely determined, in the following sense: if \( n', m' \in \mathbb{N}, \) with \( m' \leq n', \) and \( y_1, \ldots, y_{n'} \in M \) such that:
\[
M = Ry_1 \oplus \ldots \oplus Ry_{m'} \oplus Ry_{m'+1} \ldots \oplus Ry_n', \quad (D')
\]
and \( o(y_i) =: e_i \) satisfy : \( e_i \in R^\circ, \forall i \in \{1, \ldots, m'\} \) and \( e_1|e_2|\ldots|e_{m'}; e_{m'+1} = \ldots = e_{n'} = 0, \) then
\[
m = m', n = n' \quad \text{and} \quad d_i \sim e_i, \forall i \in \{1, \ldots, n\}.
\]

The “orders” \( o(x_i) \) are called the invariant factors of the module \( M. \)

**Proof.** The existence part: if \( S \subseteq M \) is finite and generates \( M, \) then there is an isomorphism \( \psi : R^S/F \rightarrow M, \) where \( F \) is the kernel of \( \varphi, \) given by \( \varphi(e_s) = s, \forall s \in S \) \((e_s)_{s \in S} \) is the canonical basis of \( R^S) \). Let \( n = |S|, E = R^S. \) Theorem 1.2 provides us with a basis \( e = (e_1, \ldots, e_n) \) in \( E \) and a basis \( f = (f_1, \ldots, f_m) \) in \( F \) (with \( m \leq n \)) such that \( f_i = d_ie_i, \) with \( d_i \in R, d_i \neq 0 \) and \( d_1|d_2|\ldots|d_m. \) There exists \( k \in \mathbb{N} \) \((0 \leq k \leq m) \) such that \( d_1, \ldots, d_k \in U(R) \) and \( d_{k+1}, \ldots, d_m \not\in U(R). \) Then \( Rf_i = Rd_ie_i = Re_i, \) \( 1 \leq i \leq k, \) and we can write (applying lemma 2.2):
\[
M \cong E/F \cong \frac{Re_1}{Re_1} \oplus \ldots \oplus \frac{Re_k}{Re_k} \oplus \frac{Re_{k+1}}{Rf_{k+1}} \oplus \ldots \oplus \frac{Re_m}{Rf_m} \oplus Re_{m+1} \oplus \ldots \oplus Re_n
\]
\[
= \frac{Re_{k+1}}{Rf_{k+1}} \oplus \ldots \oplus \frac{Re_m}{Rf_m} \oplus Re_{m+1} \oplus \ldots \oplus Re_n
\]
\[
\cong \frac{Re_{k+1}}{Rf_{k+1}} \oplus \ldots \oplus \frac{Re_m}{Rf_m} \oplus Re_{m+1} \oplus \ldots \oplus Re_n
\]

Thus, the structure of \( M \) is determined by \( d_{k+1}, \ldots, d_m, e_{k+1}, \ldots, e_n, f_{k+1}, \ldots, f_m. \) Changing notations if necessary, we may suppose from the beginning that \( d_1, \ldots, d_m \) are non invertible. Let \( \psi(e_i + F) = x_i \in M, \) \( 1 \leq i \leq n. \)

We have \( M = Rx_1 \oplus \ldots \oplus Rx_n \) because \( \{e_i + F | 1 \leq i \leq n\} \) generates \( E/F, \) so \( \{x_i | 1 \leq i \leq n\} \) generates \( M. \) Besides,
III.2 Finitely generated modules over a principal ideal domain

\[ E/F = R(e_1 + F) \oplus \ldots \oplus R(e_n + F), \]
since \( \sum_{1 \leq i \leq n} r_i(e_i + F) = 0 + F \iff \exists s_i \in R \text{ such that } \sum_{1 \leq i \leq m} r_i e_i = \sum_{1 \leq i \leq m} s_id_i e_i \iff r_i = s_id_i, \) for any \( 1 \leq i \leq m \) and \( r_i = 0, \) for \( m + 1 \leq i \leq n \) \( \iff r_i(e_i + F) = 0 + F, \) \( 1 \leq i \leq n. \)

Also, \( \o(x_i) = o(e_i + F) = d_i, \) if \( 1 \leq i \leq m \) and \( o(x_i) = 0 \) if \( m + 1 \leq i \leq n. \) Indeed, \( O_r(x_i) = 0 \iff r_{e_i} \in F \iff \exists s_i \in R \text{ such that } \sum_{1 \leq i \leq m} r_{e_i} = \sum_{1 \leq i \leq m} s_id_i, \) so \( r = 0 \) if \( m + 1 \leq i \leq n \) and \( r = s_id_i \) if \( 1 \leq i \leq m. \)

This finishes the proof of existence of a decomposition \((D)\) of \( M.\)

For the uniqueness part, we need some “invariants”.

2.4 Definitions. Let \( R \) be a domain and let \( M \) be an \( R \)-module.

a) Define \( t(M) := \{x \in M \mid \exists r \in R \setminus \{0\} \text{ with } rx = 0\}. \) \( t(M) \) is called the torsion submodule of \( M.\) The elements of \( t(M) \) are called torsion elements. If \( M = t(M), \) \( M \) is called a torsion module; if \( t(M) = 0, \) \( M \) is called a torsion-free module.

b) Let \( p \in R \) be a prime element. Define \( t_p(M) := \{x \in M \mid \exists n \in \mathbb{N} \text{ cu } p^n x = 0\} \) (called the \( p \)-torsion submodule of \( M \) or the \( p \)-submodule of \( M).\)

If \( d \in R \) and \( k \in \mathbb{N}, \) \( p^k \| d \) means \( p^k | d \) and \( p^{k+1} \| d \) (if \( R \) is a UFD, \( p^k \| d \) \( \iff k \) is the exponent of \( p \) in the prime factor decomposition of \( d).\)

c) If \( a \in R, \) let \( z_a(M) := \{x \in M \mid ax = 0\} \) (called the annihilator of \( a \) in \( M), \) also denoted sometimes by \( \text{Ann}_M(a) \) or \( r_M(a)).\)

d) \( \text{Ann}_R(M) := \{r \in R \mid rx = 0, \forall x \in M\} \) is called the annihilator of the module \( M.\)

We need to collect the basic proprieties of these invariants.

2.5 Proposition. Let \( R \) be a domain and let \( M \) be an \( R \)-module.

a) \( t(M) \) is a submodule of \( M \) and \( t(M) / t(M) = 0; \) if \( M \cong R \) \( N, \) then \( t(M) \cong t(N).\)

b) If \( M = \bigoplus_{i \in I} M_i, \) then \( t(\bigoplus_{i \in I} M_i) = \bigoplus_{i \in I} t(M_i).\)
c) Let $R$ be a PID. If $M$ is finitely generated, then $t(M)$ is a direct summand in $M$ and there exists $L \leq R M$, free, such that $M = t(M) \oplus L$. In particular, a finitely generated torsion-free module is free.

**Proof.** a) Let $x, y \in t(M)$ and let $r, s \in R$. Then there exist $a, b \in R$, nonzero, such that $ax = by = 0$. We have $ab \neq 0$ and $ab(rx + sy) = 0$, so $rx + sy \in t(M)$.

Let $x + t(M) \in t(M/t(M))$: there exists $0 \neq a \in R$ with $ax \in t(M)$. Then there exists $0 \neq b \in R$ with $bax = 0$, so $x \in t(M)$ (since $ba \neq 0$). Thus, $x + t(M) = 0 + t(M)$.

b) Let $(x_i)_{i \in I}$ have finite support, with $x_i \in M_i, \forall i \in I$, such that $\sum_{i \in I} x_i := x \in t(M)$. Then there exists $0 \neq a \in R$ such that $ax = \sum_{i \in I} ax_i = 0$. From $M = \bigoplus_{i \in I} M_i$ it follows that $ax_i = 0$, i.e. $x_i \in t(M_i), \forall i \in I$. So, $t(M) \subseteq \bigoplus_{i \in I} t(M_i)$.

If $x_i \in t(M_i)$ such that $J := \text{supp}((x_i)_{i \in I})$ is finite, there is some family $(r_j)_{j \in J}$, with $0 \neq r_j \in R$ and $r_jx_j = 0, \forall j \in J$. Let $r := \prod_{j \in J} r_j$ (well defined, since $J$ is finite). Obviously, $r \neq 0$ and $r\sum_{i \in I} x_i = 0$. This shows that $\bigoplus_{i \in I} t(M_i) \subseteq t(M)$.

c) Take a decomposition $(D)$ of $M$. We show that $t(M) = Rx_1 \oplus \ldots \oplus Rx_m$. Evidently, $x_1, \ldots, x_m \in t(M)$, since $d_ix_i = 0$ and $d_i \neq 0$, so $Rx_1 \oplus \ldots \oplus Rx_m \subseteq t(M)$. If $x \in t(M)$, there is some $0 \neq a \in R$ and $r_1, \ldots, r_n \in R$ such that $x = r_1x_1 + \ldots + r_mx_m + \ldots + r_nx_n$ and $ax = ar_1x_1 + \ldots + ar_mx_m + \ldots + ar_nx_n = 0$. Because the sum of the submodules $Rx_i$ is direct, $ar_i x_i = 0, i = 1, \ldots, n$. If $i > m$, then $ar_i \in \text{Ann}_R(x_i) = 0$, so $r_i = 0$. We deduce that $x \in Rx_1 \oplus \ldots \oplus Rx_m$.

It is now clear that, denoting by $L$ the free module $\bigoplus_{i \geq m} Rx_i$, $M = t(M) \oplus L$. $\square$

2.6 **Proposition.** Let $R$ be a PID, let $M$ be an $R$-module and $p \in R$ a prime element.

a) $t_p(M)$ is a submodule in $M$ and $t_p(M/t_p(M)) = 0$; if $M \cong_R N$, then $t_p(M) \cong t_p(N)$.

b) If $M = \bigoplus_{i \in I} M_i$, then $t_p(\bigoplus_{i \in I} M_i) = \bigoplus_{i \in I} t_p(M_i)$. 

III. Finitely generated modules over principal ideal domains
III.2 Finitely generated modules over a principal ideal domain

particular, if \( x \in M \) and \( o(x) = d \), then \( t_p(Rx) \cong R/Rp^k \); if \( p|d \), then \( t_p(Rx) = 0 \).

**Proof.** \( a \) and \( b \) have similar proofs with 2.5 and are proposed as an exercise.

\( c \) Let \( b \in R \) such that \( d = p^k b \). We claim that \( t_p(R/Rd) = Rb/Rd \). Indeed, let \( r + Rd \in t_p(R/Rd) \). There exist \( s \in \mathbb{N} \) such that \( p^s r \in Rd \), i.e. \( p^s r = dc = p^k bc \), with \( c \in R \). So \( b|p^s r \) and \((b, p) = 1\), which imply \( b|r \).

So, \( r \in Rb \) and \( t_p(R/Rd) \subseteq Rb/Rd \). The other inclusion is obvious.

We have \( R/Rp^k \cong Rb/Rd \) by the isomorphism theorem applied to \( \varphi : R \to Rb/Rd, \varphi(r) = rb + Rd, \forall r \in R \).

2.7 **Proposition.** Let \( M \) be an \( R \)-module and let \( r \in R \).

\( a \) \( z_r(M) \) is a submodule in \( M \); if \( M \cong_R N \), then \( z_r(M) \cong_R z_r(N) \).

\( b \) If \( M = \bigoplus_{i \in I} M_i \), then \( z_r(M) = \bigoplus_{i \in I} z_r(M_i) \).

\( c \) If \( R \) is a PID, \( p \in R \) is a prime element and \( d \in R \), then \( z_p(R/Rd) = 0 \) if \( p|d \) and \( z_p(R/Rd) \cong R/Rp \) if \( p|d \). In particular, if \( x \in M \), then \( z_p(Rx) = 0 \) if \( p|o(x) \) and \( z_p(Rx) \cong R/Rp \) if \( p|o(x) \).

**Proof.** \( a \) Exercise.

\( b \) Let \( x = \sum_{i \in I} x_i \), with \( x_i \in M_i, \forall i \in I \). We have \( rx = 0 \iff \sum_{i \in I} rx_i = 0 \iff rx_i = 0, \forall i \in I \) (since \( rx_i \in M_i \), and the sum is direct) \iff \( x_i \in z_r(M_i), \forall i \in I \).

\( c \) Let \( p \nmid d \) and \( r \in R \). We have \( p(r + Rd) = 0 + Rd \iff pr \in Rd \iff d | pr \). Since \((d, p) = 1\), this implies \( d | r \), so \( r + Rd = 0 \).

Let \( p|d \) and \( a \in R \) with \( d = pa \). For any \( r \in R \), \( pr \in Rd = Rpa \iff r \in Ra \). So, \( z_p(R/Rd) = Ra/Rd \). Consider the surjective homomorphism \( \varphi : R \to Ra/Rd, \varphi(r) = ra + Rd, \forall r \in R \). We have \( \text{Ker}\varphi = Rp \), so \( Ra/Rd \cong R/Rp \).

We can proceed now to prove the **uniqueness** part in 2.3. The idea is to apply the invariants \( t, z_p, t_p \) to the decompositions \((D)\) and \((D')\).

Since \( t(M) = Rx_1 \oplus \ldots \oplus Rx_m = Ry_1 \oplus \ldots \oplus Ry_m \) (see 2.5.c), we get:
$M/t(M) \cong Rx_{m+1} \oplus \ldots \oplus Rx_n \cong Ry_{m'+1} \oplus \ldots \oplus Ry_{n'}$.

But $\{x_{m+1}, \ldots, x_n\}$ and $\{y_{m'+1}, \ldots, y_{n'}\}$ are bases in the free module $M/t(M)$, so they have the same number of elements: $n - m = n' - m'$.

It remains to prove that $m = m'$ and $d_i \sim e_i$, $1 \leq i \leq m$. To this end, we suppose in the sequel that

$$M = t(M) = Rx_1 \oplus \ldots \oplus Rx_m = Ry_1 \oplus \ldots \oplus Ry_{m'}.$$  

Since $Rx_i \cong R/Rd_i$, $Ry_i \cong R/Re_i$, we have to prove that, if

$$M \cong R/Rd_1 \oplus \ldots \oplus R/Rd_m, \text{ with } d_1|d_2|\ldots|d_m, d_i \in R^\circ, 1 \leq i \leq m,$$  

$$M \cong R/Re_1 \oplus \ldots \oplus R/Re_{m'}, \text{ with } e_1|e_2|\ldots|e_{m'}, e_i \in R^\circ, 1 \leq i \leq m',$$  

then $m = m'$ and $d_i \sim e_i$, $1 \leq i \leq m$.

Note first that $d_m \sim e_{m'}$. Indeed, $\text{Ann}_R(M) = \{r \in R | \text{ } rx = 0, \forall x \in M\}$ is an ideal in $R$; by (*) $\text{Ann}_R(M) = Rd_m$ (prove!). By (**), $\text{Ann}_R(M) = Re_{m'}$, so $d_m \sim e_{m'}$.

Let $p$ be a prime divisor of $d_i$. Then $p|d_i$, $1 \leq i \leq m$. Consider $z_p(M)$. By 2.7 and (*), we can write the $R$-module isomorphisms:

$$z_p(M) \cong z_p(Rx_1 \oplus \ldots \oplus Rx_m) \cong z_p(Rx_1) \oplus \ldots \oplus z_p(Rx_m) \cong R/Rp \oplus \ldots \oplus R/Rp$$  

($m$ terms).

(we used that $p|d_i$, $1 \leq i \leq m$). Use now (**); let $k$ be the number of those indices $i$, with $1 \leq i \leq m'$, such that $p$ divides $e_i$. We have:

$$z_p(M) \cong z_p(Ry_1) \oplus \ldots \oplus z_p(Ry_{m'}) \cong R/Rp \oplus \ldots \oplus R/Rp$$  

($k$ terms).

So, $(R/Rp)^m \cong (R/Rp)^k$ ($R$-module isomorphism). It is clear that this is also an $R/Rp$-module isomorphism. But $R/Rp$ is a field ($p$ is prime and $R$ is a PID), so $m = k$, since two isomorphic $R/Rp$-vector spaces have the same dimension. Of course, $k \leq m'$, so $m \leq m'$. By symmetry, $m' \leq m$, so $m = m'$.

Let us show that $d_i \sim e_i$, $1 \leq i \leq m$. Let $p$ be an arbitrary prime factor of $d_m$ and let $\alpha_i$ (respectively $\beta_i$) be the exponent of $p$ in the decomposition of $d_i$ (respectively $e_i$), $1 \leq i \leq m$. Since $d_1|d_2|\ldots|d_m$, we have $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_m$ (and also $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_m$). It is sufficient to show that $\alpha_i = \beta_i$, $1 \leq i \leq m$. Suppose that is not so: then there exists $j$, 

...
1 \leq j \leq m$, minimal with the property that $\alpha_i \neq \beta_j$ (thus $\alpha_i = \beta_i$ if $i < j$).

Say, for instance, that $\alpha_j < \beta_j$. Apply $t_p$. By (*) and 2.6, we have:

$$t_p(M) \cong t_p(Rx_1) \oplus \ldots \oplus t_p(Rx_m) \cong R/Rp^{\alpha_1} \oplus \ldots \oplus R/Rp^{\alpha_m}$$

By (**), $t_p(M) \cong t_p(Ry_1) \oplus \ldots \oplus t_p(Ry_m) \cong R/Rp^{\beta_1} \oplus \ldots \oplus R/Rp^{\beta_m}$.

“Multiply” $t_p(M)$ by $p^{\alpha_j}$ (i.e., consider the submodule $p^{\alpha_j}t_p(M)$). We obtain from (*):

$$p^{\alpha_j}t_p(M) \cong p^{\alpha_j}(R/Rp^{\alpha_1}) \oplus \ldots \oplus p^{\alpha_j}(R/Rp^{\alpha_m})$$

$$\cong R/Rp^{\alpha_{j+1}-\alpha_j} \oplus \ldots \oplus R/Rp^{\alpha_m-\alpha_j}$$

We have used the proprieties (whose proof is immediate): if $M = \bigoplus_{i \in I} M_i$ and $r \in R$, then $rM = \bigoplus_{i \in I} rM_i$; $p^{\alpha}(R/Rp^{\beta}) \cong 0$ if $\beta \leq \alpha$; $p^{\alpha}(R/Rp^{\beta}) \cong R/Rp^{\beta-\alpha}$ if $\beta > \alpha$.

From (**), with a similar argument, we have:

$$p^{\alpha_j}t_p(M) \cong p^{\alpha_j}(R/Rp^{\beta_1}) \oplus \ldots \oplus p^{\alpha_j}(R/Rp^{\beta_n})$$

$$\cong R/Rp^{\beta_{j+1}-\alpha_j} \oplus \ldots \oplus R/Rp^{\beta_m-\alpha_j}$$

Therefore,

$$R/Rp^{\alpha_{j+1}-\alpha_j} \oplus \ldots \oplus R/Rp^{\alpha_m-\alpha_j} \cong R/Rp^{\beta_{j+1}-\alpha_j} \oplus \ldots \oplus R/Rp^{\beta_m-\alpha_j}$$

These are decompositions of the type (*), as one easily sees. Let $k$ be the number of indices $i$ for which $\alpha_i > \alpha_j$ (evidently, $0 \leq k \leq m - j$). In the left hand side we have $k$ nonzero terms and in the right hand side there are $m - j + 1$ nonzero terms. The first part of the proof shows that $k = m - j + 1$, contradicting $k \leq m - j$. Thus, we must have $\alpha_i = \beta_i$, $1 \leq i \leq m$.

\[\square\]

2.8 Remark. An $R$-module $M$ can have several decompositions of type (D). But the sequence of ideals

$$\text{Ann}_R(x_1) \supseteq \text{Ann}_R(x_2) \supseteq \ldots \supseteq \text{Ann}_R(x_n)$$

is uniquely determined. The generators (uniquely determined up to an association in divisibility) $d_1$, $d_2$, ..., $d_n \in R$ of these ideals are called the invariant factors of the $R$-module $M$. Sometimes this name is
given to the sequence of ideals above. If in the principal ideal domain $R$ there is a natural way to choose a generator of an ideal, the invariant factors are the “natural” generators of the annihilator ideals above. For example, in $\mathbb{Z}$ the positive generator is chosen, in $K[X]$ ($K$ a field) the monic polynomial that generates the ideal is chosen.

2.9 Example. Let $\mathbb{Z}_2 = \{ \hat{0}, \hat{1} \}$, $\mathbb{Z}_6 = \{ \bar{0}, \bar{1}, \ldots, \bar{5} \}$ and the $\mathbb{Z}$-module $\mathbb{Z}_2 \times \mathbb{Z}_6$. We have the following decompositions:

$$\mathbb{Z}_2 \times \mathbb{Z}_6 = \mathbb{Z}x_1 \oplus \mathbb{Z}x_2 = \mathbb{Z}y_1 \oplus \mathbb{Z}y_2,$$

where $x_1 = (\hat{1}, \hat{0})$, $x_2 = (\hat{0}, \bar{1})$, $y_1 = (\hat{1}, \bar{3})$, $y_2 = (\hat{0}, \bar{5})$ (check this!). Of course, these are distinct decompositions, but $\text{Ann}_R(x_1) = \text{Ann}_R(y_1) = 2\mathbb{Z}$ and $\text{Ann}_R(x_2) = \text{Ann}_R(y_2) = 6\mathbb{Z}$. The sequence of invariant factors of the $\mathbb{Z}$-module $\mathbb{Z}_2 \times \mathbb{Z}_6$ is: 6, 2.

Exercises

1. Give an example of a finitely generated $\mathbb{Z}$-module such that its torsion submodule has at least two complements. (cf. Prop. 2.5.)

2. Let $R$ be a domain and let $u : M \to N$ be an $R$-module homomorphism. Then $u(t(M)) \subseteq t(N)$. State and prove a similar propriety for $t_p(M)$ ($R$ is a PID and $p$ is a prime element in $R$) and $z_r(M)$ (where $r \in R$).

3. Assume $R$ is a domain, $M$ is an $R$-module and $L \leq_R M$. Then $M/L$ is torsion-free if and only if $t(M) \subseteq L$.

4. This exercise gives an example of a $\mathbb{Z}$-module $M$ such that $t(M)$ is not a direct summand in $M$. (Thus, by 2.5.c), $M$ cannot be finitely generated). Let $P = \{ p_n \mid n \geq 1 \}$ be the set of prime natural numbers and let $M := \prod_{p \in P} \mathbb{Z}_p$. Show that:
III.2 Finitely generated modules over a principal ideal domain

a) \( t(M) = \bigoplus_{p \in P} \mathbb{Z}_p \) (\( = \{(a_p)_{p \in P} | \forall p \in P, \ a_p \in \mathbb{Z}_p \) and \( \text{supp}((a_p)_p) \) finite\)).

b) \( \forall n \in \mathbb{Z}, n \neq 0, n \cdot (M/t(M)) = M/t(M) \).

c) \( \forall x \in M, \exists p \in P \) such that \( x \notin pM \).

d) If \( M = t(M) \bigoplus S, \) with \( S \leq R M \), then \( S \cong M/t(M) \).

e) \( t(M) \) is not a direct summand in \( M \).

5. Let \( p \) be a prime natural number and let \( G = \mathbb{Z}_p \times \mathbb{Z}_p \). Prove that the subgroups of \( G \) coincide with the \( \mathbb{Z}_p \)-vector subspaces of \( G \). How many vector subspaces of dimension 1 are there in \( G \)? In how many ways can \( \mathbb{Z} G \) be written as a direct sum of two proper submodules? (Hint: the sum of any two nonzero distinct subspaces is direct, equal to \( G \)).

6. State a structure theorem for finitely generated Abelian groups.

7. Give an example of an Abelian group \( G \) that is neither cyclic, nor a direct sum of cyclic groups. (Hint: \( G \) cannot be finitely generated. An example is \( \mathbb{Q} \)).

8. Determine all Abelian groups with 100 elements.

9. Let \( G \) be a finite Abelian group whose order is squarefree (not divisible with the square of any prime). Then \( G \) is cyclic.

10. If \( G \) is a finite Abelian group and \( n \) is the exponent of \( G \) (\( = \text{GCD} \) of the orders of the elements of \( G \)), then \( G \) contains an element of order \( n \). (Hint: use the theorem of invariant factors). Give an example of a group \( G \) such that \( n \) is a proper divisor of the order of \( G \).

11. Let \( G \) be a finite Abelian group. Show that, for any divisor \( d \) of \( |G| \), there exists a subgroup of \( G \) having \( d \) elements.

12. For \( n \in \mathbb{N}^* \), let \( g_n \) denote the number of isomorphism types of Abelian groups having \( n \) elements. For what \( n \in \mathbb{N}^* \) we have \( g_n = 1 \)? Determine \( \{n \in \mathbb{N} | n \leq 100, g_n \geq 4 \} \).

13. Using the structure theorem of finite Abelian groups, show that any finite subgroup \( G \) of the multiplicative group \( K^* \) (where \( K \) is a
field) is cyclic. (*Hint.* If $G$ has more than one invariant factor and $d$ is the greatest invariant factor, then any element of $G$ is a root of a $X^d - 1$ and $d < |G|$).

14. Determine the finitely generated Abelian groups with the property that their lattice of subgroups is a chain (= totally ordered with respect to inclusion).

### III.3 Indecomposable finitely generated modules

Theorem 2.3 says that any finitely generated $R$-module (if $R$ is a PID) is a direct sum of cyclic submodules. Can such a decomposition be refined? In other words, can we decompose further these submodules as direct sums of proper submodules?

**3.1 Definition.** Let $R$ be a ring (not necessarily commutative). An $R$-module $M$ is called indecomposable if $M \neq 0$ and $M$ has no proper direct summands (if $L, N \leq_R M$ are such that $M = L \oplus N$, then $L = 0$ or $N = 0$). An $R$-module that is not indecomposable is called decomposable.

**3.2 Examples.** a) The $\mathbb{Z}$-module $\mathbb{Z}_6$ is decomposable: $\mathbb{Z}_6 = 2\mathbb{Z}_6 \oplus 3\mathbb{Z}_6$.

b) If $R M$ is isomorphic to a direct product of modules of the type $A \times B$, with $A, B$ nonzero, then $M$ is decomposable.

b) If $K$ is a field, a $K$-vector space $V$ is decomposable if and only if $\dim V > 1$. (Besides, any proper subspace of a vector space is a direct summand).

c) The $\mathbb{Z}$-module $\mathbb{Z}_2$ is indecomposable (it has no proper submodules altogether).
III.3 Indecomposable finitely generated modules

3.3 Proposition. If the lattice of submodules of the R-module M is a chain (it is totally ordered) with respect to inclusion, then M is indecomposable.

Proof. Suppose $M = A \oplus B$, with $A, B \leq_R M$. But the lattice of submodules of M is a chain, so $A \subseteq B$ or $B \subseteq A$. Thus, $A \cap B = 0$ implies $A = 0$ or $B = 0$. □

3.4 Corollary. Let R be a PID, $p \in R$ a prime element and $k \in \mathbb{N}$. Then the cyclic module $R/Rp^k$ is indecomposable.

Proof. Suppose $k \geq 1$. It is enough to check that the lattice of ideals of R that include $Rp^k$ is a chain (see II.2.3). But the ideal I includes $Rp^k$ if and only if $I = Ra$ (I is principal), where $a \in R$, $a|p^k$. The divisors of $p^k$ are $p^t$, $0 \leq t \leq k$, so $a \sim p^t$ for some $t \leq k$. This says that the ideals that include $Rp^k$ are chained: $Rp^k \subseteq Rp^{k-1} \subseteq \ldots \subseteq Rp^2 \subseteq Rp \subseteq R$.

If $k = 0$, we have to show that $_R R$ is indecomposable. This is true for an arbitrary domain: the intersection of any two nonzero submodules (= ideals) I, J of R is nonzero: if $a \in I$, $b \in J$ are nonzero, then $0 \neq ab \in I \cap J$. □

If R is a PID, the cyclic modules (isomorphic to) $R/Rp^k$, with $p$ prime in R and $k \geq 0$, are all finitely generated indecomposable R-modules. This follows from the following result, valid in any commutative ring.

3.5 Theorem. (Chinese remainder theorem) Let R be a commutative ring, $n \geq 2$ and $I_1, \ldots, I_n$ ideals of R.

a) If $I_i + I_j = R$ for $i \neq j$, then the product $I_1 \cdots I_n$ is equal to the intersection $I_1 \cap \ldots \cap I_n$ and there is a natural isomorphism of rings (and of R-modules) $\eta$:

---

3 The ideals $I_i$ and $I_j$ are called in this case comaximal. For example, the ideals $\mathbb{Z}a$ and $\mathbb{Z}b$ of $\mathbb{Z}$ are comaximal if and only if $a$ and $b$ are coprime.
III. Finitely generated modules over principal ideal domains

\[
\frac{R}{I_1 \cap \ldots \cap I_n} = \frac{R}{I_1 \times \ldots \times I_n},
\]

\[
\eta(r + I_1 \cap \ldots \cap I_n) = (r + I_1, \ldots, r + I_n), \forall r \in R.
\]

b) Conversely, if the homomorphism \( \varphi : R \to \frac{R}{I_1 \times \ldots \times I_n} \), \( \varphi(r) = (r + I_1, \ldots, r + I_n), \forall r \in R \) is surjective (inducing an isomorphism \( \frac{R}{I_1 \cap \ldots \cap I_n} \cong \frac{R}{I_1} \times \ldots \times \frac{R}{I_n} \) as above), then \( I_i \) and \( I_j \) are comaximal ideals, for any \( i \neq j \).

**Proof.** a) We prove by induction on \( n \) that \( I_1 \cdot \ldots \cdot I_n = I_1 \cap \ldots \cap I_n \) and that \( \eta \) is an isomorphism. For \( n = 2 \), \( I_1 + I_2 = R \) implies the existence of \( x \in I_1, \ y \in I_2 \) such that \( x + y = 1 \). Let \( z \in I_1 \cap I_2 \). Then \( z = z \cdot 1 = zx + zy \), with \( zx, zy \in I_1 : I_2 \), so \( I_1 \cap I_2 \subseteq I_1 I_2 \). Thus, \( I_1 \cap I_2 = I_1 I_2 \).

Let \( \varphi : R \to \frac{R}{I_1} \times \frac{R}{I_2} \), \( \varphi(r) = (r + I_1, r + I_2), \forall r \in R \). \( \varphi \) is a ring (and an \( R \)-module) homomorphism (it is the direct product of the canonical surjections \( R \to R/I_j \)). We have:

\[
\text{Ker} \varphi = \{ r \in R | (r + I_1, r + I_2) = (0 + I_1, 0 + I_2) \} = I_1 \cap I_2
\]

The isomorphism theorem implies that \( R/I_1 \cap I_2 \cong \text{Im} \varphi \) and the isomorphism is precisely \( \eta \). Let us prove that \( \varphi \) is surjective (which will finish the proof). Let \( (r_1 + I_1, r_2 + I_2) \in \frac{R}{I_1} \times \frac{R}{I_2} \). We must exhibit \( r \in R \) with \( r - r_1 \in I_1, r - r_2 \in I_2 \). Such an element is \( r = r_1 y + r_2 x \). Indeed,

\[
r - r_1 = r_1 y + r_2 x - r_1 x = (r_2 - r_1)x \in I_1.
\]

\[\text{Recall that the product } IJ \text{ of two ideals } I \text{ and } J \text{ is the ideal generated by all the products } ij, i \in I, j \in J. \text{ The product of ideals is associative and always } IJ \subseteq I \cap J.\]
Similarly, \( r - r_2 \in I_2 \).

Suppose now that for any \( k < n \) and any ideals \( I_1, \ldots, I_k \), pairwise comaximal, we have \( I_1 \cdots I_k = I_1 \cap \cdots \cap I_k \) and \( \eta \) is an isomorphism. Take \( n \) pairwise comaximal ideals \( I_1, \ldots, I_n \). Since \( I_j + I_n = R \), \( 1 \leq j \leq n - 1 \), there exist \( a_j \in I_j \), \( b_j \in I_n \), such that \( a_j + b_j = 1 \). Multiply these \( n - 1 \) equalities:

\[
\prod_{j=1}^{n-1} (a_j + b_j) = a_1 \cdots a_{n-1} + b = 1, \quad \text{where } b \in I_n, a_1 \cdots a_{n-1} \in I_1 \cdots I_{n-1}.
\]

So, \( I_1 \cdots I_{n-1} + I_n = R \). Applying the case \( n = 2 \) to the comaximal ideals \( I_1 \cdots I_{n-1} \) and \( I_n \),

\[
I_1 \cdots I_{n-1} \cdot I_n = (I_1 \cdots I_{n-1}) \cap I_n = (I_1 \cap \cdots \cap I_{n-1}) \cap I_n
\]

We also used the induction hypothesis \( I_1 \cdots I_{n-1} = I_1 \cap \cdots \cap I_{n-1} \). Also, case \( n = 2 \) says that

\[
\begin{align*}
R & \cong I_1 \cdots I_{n-1} = I_1 \cap \cdots \cap I_{n-1} \times I_n, \\
(1 + I_1, 0 + I_n) & \mapsto (r + I_1 \cdots I_{n-1}, r + I_n), \quad \forall r \in R.
\end{align*}
\]

On the other hand, by induction, we have the isomorphism:

\[
\begin{align*}
R & \cong I_1 \cdots I_{n-1} = I_1 \cap \cdots \cap I_{n-1}, \\
(1 + I_1, \ldots, 0 + I_n) & \mapsto (r + I_1, \ldots, r + I_{n-1}), \quad \forall r \in R.
\end{align*}
\]

Combining these isomorphisms, we obtain the result.

\( b) \) We prove that \( I_1 \) and \( I_2 \) are comaximal. Let \( (1 + I_1, 0 + I_2, \ldots, 0 + I_n) \in R \times \cdots \times R \). There exists \( y \in R \) such that \( (y + I_1, y + I_2, \ldots, y + I_n) = (1 + I_1, 0 + I_2, \ldots, 0 + I_n) \), i.e. \( y \in I_2 \) and \( y - 1 = x \in I_1 \). So, \( 1 = -x + y \in I_1 + I_2 \), which means \( I_1 + I_2 = R \).

\[\square\]

**3.6 Corollary.** \( a) \) Let \( R \) be a PID and let \( a_1, \ldots, a_n \in R \). If \( (a_i, a_j) = 1, \; \forall i \neq j \), then:
III. Finitely generated modules over principal ideal domains

$$\frac{R}{Ra_{1} \ldots a_{n}} \cong \frac{R}{Ra_{1}} \times \ldots \times \frac{R}{Ra_{n}}$$

$$r + Ra_{1} \ldots a_{n} \mapsto (r + Ra_{1}, \ldots, r + Ra_{n}), \forall r \in R.$$

b) If \(M\) is a cyclic \(R\)-module \(M = Rx (x \in M)\), with \(o(x) = d = a_{1} \cdot \ldots \cdot a_{n} \in R^{\circ}\) and \((a_{i}, a_{j}) = 1, \forall i \neq j\), then there exist \(x_{i} \in M, 1 \leq i \leq m\), such that \(o(x_{i}) = a_{i}\) and

$$M = Rx = Rx_{1} \oplus \ldots \oplus Rx_{m}$$

c) Any cyclic \(R\)-module can be written as a direct sum of indecomposable submodules.

**Proof.**
a) If \(a, b \in R\), then \((a, b) = 1\) if and only if \(Ra\) and \(Rb\) are co-maximal. Indeed, the ideal generated by \(\text{GCD}(a, b)\) is \(Ra + Rb\). So, \((a, b) = 1 \iff Ra + Rb = R\). Apply now the Chinese remainder theorem for \(Ra_{1}, \ldots, Ra_{n}\).

b) We have \(M \cong R/Rd\). By a), we have \(R/Ra_{1} \times \ldots \times R/Ra_{n} \cong R/Rd\).

So, there is an isomorphism \(\varphi : R/Ra_{1} \times \ldots \times R/Ra_{n} \to M\). Let \(y_{i} := (0 + Ra_{1}, \ldots, 1 + Ra_{i}, \ldots, 0 + Ra_{n})\) and \(x_{i} := \varphi(y_{i})\). Obviously, \(R/Ra_{1} \times \ldots \times R/Ra_{n} = Ry_{1} \oplus \ldots \oplus Ry_{n}\), so, (applying \(\varphi\)) \(M = Rx_{1} \oplus \ldots \oplus Rx_{n}\).

Also, \(o(x_{i}) = o(y_{i}) = a_{i}, 1 \leq i \leq n\).

c) This follows from a): let \(M = Rx, x \in M\) and let \(d = o(x)\). If \(d = 0\), then \(M \cong R\) is indecomposable. If \(d \neq 0\), let \(d = p_{1}^{k_{1}} \ldots p_{t}^{k_{t}}\) the prime factor decomposition of \(d\) (where \(p_{1}, \ldots, p_{t}\) are distinct primes in \(R\)). Clearly, \(p_{i}^{k_{i}}\) and \(p_{j}^{k_{j}}\) are coprime if \(i \neq j\); applying b), there exist \(x_{i} \in M\) such that \(M = Rx_{1} \oplus \ldots \oplus Rx_{t}\), with \(o(x_{i}) = p_{i}^{k_{i}}\). So \(Rx_{i}\) is indecomposable, being isomorphic to \(R/Rp_{i}^{k_{i}}\).

3.7 Corollary. Let \(R\) be a PID and let \(M\) be a finitely generated \(R\)-module. Then:

- \(M\) is indecomposable if and only if: either \(M\) is cyclic, isomorphic to \(R\) (in this case \(M\) is torsion-free) or \(M\) is cyclic, isomorphic to some
III.3 Indecomposable finitely generated modules

$R/Rp^k$, where $p \in R$ is prime and $k \in \mathbb{N}^*$ (in this case $M$ is a torsion module).

**Proof.** The proof of the “if” part is easy: we saw that the modules $R/Rp^k$ are indecomposable.

Let $M$ be indecomposable finitely generated. There exists a decomposition $(D)$, as in theorem 2.3. Keeping the notations in 2.3, we see that $M$ is indecomposable only if $m = n = 1$ or $m = 0$ and $n = 1$. If $m = 0$, $n = 1$, then $M$ is free of rank 1, and thus isomorphic to $R$.

If $m = n = 1$, then $M = Rx$, with $o(x) = d$. So, $M \cong R/Rd$. Let $d = p_1^{k_1} \ldots p_t^{k_t}$ be the prime decomposition of $d$ ($p_1, \ldots, p_t$ are distinct primes in $R$). We claim that $t = 1$. If $t > 1$, $p_i^{k_i}$ and $p_j^{k_j}$ are coprime if $i \neq j$, so $R/Rd \cong R/Rp_1^{k_1} \times \ldots \times R/Rp_t^{k_t}$, which is clearly decomposable.

3.8 Proposition. If $M$ is a finitely generated torsion module over a PID $R$, then $M$ is the direct sum of its $p$-torsion submodules.

More precisely, if $P$ is a system of representatives of the equivalence classes (with respect to association in divisibility) of the prime elements in $R$, then $\{p \in P \mid t_p(M) \neq 0\}$ is finite and $M$ is the finite direct sum:

$$M = \bigoplus_{p \in P} t_p(M)$$

**Proof.** We know that $M$ is a direct sum of cyclic submodules: $M = Rx_1 \oplus \ldots \oplus Rx_m$, $o(x_i) = d_i \in R^\circ$. From 2.6 we deduce that, if $p \in P$ and $p|d_i$, for all $1 \leq i \leq m$, then $t_p(M) = 0$. So, $\{p \in P \mid t_p(M) \neq 0\} \subseteq \{p \in P \mid \exists i, p|d_i\}$, which is finite.

If $p \in P$, then $t_p(M) \cap (\sum\{t_q(M) \mid q \in P, q \neq p\}) = 0$. Indeed, if $x$ belongs to the intersection, then $p^kx = 0$ for some $k \in \mathbb{N}$; also, $x = \sum_{q \neq p} y_q$, with $y_q \in t_q(M)$, $\forall q \in P, q \neq p$ (only a finite number of $y_q$ are nonzero). So, $\forall q \neq p$ for which $y_q \neq 0$, $\exists k_q \in \mathbb{N}$ such that $q^{k_q} y_q = 0$. Let $a = \prod_{y_q \neq 0} q^{k_q}$ (finite number of factors!). Evidently, $ax =$
a\sum_{q \neq p} y_q = 0. Also (p^k, a) = 1, since p is not associated to any prime q occurring in the decomposition of a. So, there exist u, v \in R such that up^k + va = 1. Then x = (up^k + va)x = up^k x + vax = 0.

Let x \in M. We prove that x \in \sum_{p \in P} t_p(M). Let d = o(x) \in R and d = p_1^{k_1} \cdots p_i^{k_i} be its prime decomposition. By 3.6.b), Rx = Rx_1 \oplus \cdots \oplus Rx_i, with o(x_i) = p_i^{k_i}, i.e. x_i \in t_{p_i}(M).

The invariant factor theorem and the list of finitely generated indecomposable modules allow us to formulate the following structure theorem:

3.9 Theorem. Let R be a PID and let M be a finitely generated R-module. Then M is a finite direct sum of indecomposable submodules. The following property of uniqueness holds: if

\[ M = A_1 \oplus \cdots \oplus A_m = B_1 \oplus \cdots \oplus B_n \]

are decompositions of M as direct sums of indecomposable submodules, then m = n and there exists a permutation \( \sigma \in S_n \) such that \( A_i \cong B_{\sigma(i)} \), 1 \( \leq \) i \( \leq \) m.

Proof. The existence part is proven as follows: from theorem 2.3, M is a finite direct sum of cyclic submodules, and 3.6 says that every cyclic submodule is a finite direct sum of indecomposable submodules.

As in the proof of uniqueness in theorem 2.3, note that the free module \( M/t(M) \) has rank equal to the cardinal of the set \{i \mid 1 \leq i \leq m, A_i \text{ torsion free}\} (and equal to the cardinal of the set \{j \mid 1 \leq j \leq n, B_j \text{ torsion free}\}). We may suppose then that

\[ M = t(M) = Rx_1 \oplus \cdots \oplus Rx_m = Ry_1 \oplus \cdots \oplus Ry_n, \]

with \( x_i, y_j \in M \) and \( Rx_i, Ry_j \) indecomposable (by 3.7, \( o(x_i) \), \( o(y_j) \) are prime powers in R). Let p be a prime in R. We have \( t_p(Rx_i) = 0 \) if \( p \mid o(x_i) \) and \( t_p(Rx_i) = Rx_i \) if \( o(x_i) \) is a power of p (see 2.6). Thus

\[ t_p(M) = \oplus \{Rx_i \mid 1 \leq i \leq m, p \mid o(x_i)\} = \oplus \{Ry_j \mid 1 \leq j \leq n, p \mid o(y_j)\}. \]
Relabeling if necessary, let \( \{i \mid 1 \leq i \leq m, p | o(x_i)\} =: \{1, \ldots, r\} \) and \( \{j \mid 1 \leq j \leq n, p | o(y_j)\} =: \{1, \ldots, s\} \), such that \( o(x_i) = p^{k_i} \) and \( o(y_j) = p^{l_j} \), with \( k_1 \leq \ldots \leq k_r \) and \( l_1 \leq \ldots \leq l_s \). Then

\[
t_p(M) = R x_1 \oplus \ldots \oplus R x_r = R y_1 \oplus \ldots \oplus R y_s, \quad o(x_1)|\ldots|o(x_r), \quad o(y_1)|\ldots|o(y_s).
\]

The uniqueness part of theorem 2.3 says that \( r = s \) and \( o(x_i) = o(y_i) \) (so, \( R x_i \cong R y_i \), \( 1 \leq i \leq r \)). Since \( M = \bigoplus_{p \in P} t_p(M) \) (see 3.8), the proof is finished. \( \square \)

### 3.10 Remark.
If \( M \) is a torsion module and \( R x_1 \oplus \ldots \oplus R x_m \) is a decomposition of \( M \) as a direct sum of indecomposables, then the family of ideals \( (\text{Ann}_R(x_i))_{1 \leq i \leq m} \) is independent of the chosen decomposition. The family of elements \( (o(x_i))_{1 \leq i \leq m} \) is thus uniquely determined (up to an association in divisibility and a permutation) by the module \( M \) and is called the family\(^5\) of the elementary divisors of \( M \). The elementary divisors are powers of prime elements in \( R \), by 3.7.

### 3.11 Examples.

a) The elementary divisors of the \( \mathbb{Z} \)-module \( \mathbb{Z}_6 \oplus \mathbb{Z}_{24} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_6 \oplus \mathbb{Z}_3 \) are \( (2, 2^3, 3, 3) \). Its invariant factors are \( (6, 24) \).

b) If \( G \) is an Abelian group with \( n \) elements, and \( d_1|\ldots|d_m \) is the sequence of its invariant factors, then \( G \cong \mathbb{Z}_{d_1} \oplus \ldots \oplus \mathbb{Z}_{d_m} \), so \( n = d_1 \cdot \ldots \cdot d_m \). Thus, any Abelian group of order \( n \) is perfectly determined (up to an isomorphism) by an \( m \)-uple \( (d_1, \ldots, d_m) \) of natural numbers \( (m \geq 1) \), such that: \( d_1 \geq 2, \ d_1 \cdot \ldots \cdot d_m = n \) and \( d_1|\ldots|d_m \). For example, for \( n = 60 \), the possible choices for \( (d_1, \ldots, d_m) \) are \( (60, 2, 30) \). Thus there are two types of isomorphism\(^6\) of a Abelian groups.

\(^5\) We avoid the term “set”, because in a set all elements are distinct, while the elementary divisors can occur more than once.

\(^6\) The class of all groups isomorphic to a given group \( G \) is called the type of isomorphism of \( G \). (This definition can be generalized to any type of algebraic
with 60 elements: \( \mathbb{Z}_{60} \cong \mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5 \) (the elementary divisors are \((2^2, 3, 5)\)) and \( \mathbb{Z}_2 \oplus \mathbb{Z}_{30} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5 \) (the elementary divisors are \((2, 2, 3, 5)\)).

c) If the invariant factors \((d_1, \ldots, d_m)\) are given, the elementary divisors can be obtained by decomposing in a product of powers of primes every \(d_i\) and writing down all the powers that occur, as many times that they arise.

Conversely, if the family of the elementary divisors is given, the invariant factors are obtained as follows: write a product containing (only once) all primes in the family of the elementary divisors, at the greatest power. The product obtained is \(d_m\) (the “largest” invariant factor – divisibility-wise). Erase from the family of the elementary divisors the powers written in the product and repeat the procedure with what is left. Continue until the elementary divisors are exhausted.

For example, if the family of the elementary divisors of a \(\mathbb{Z}\)-module is \((2, 2, 2^2, 3, 3^3, 5)\), following the procedure above we obtain successively: \(2^2 \cdot 3^3 \cdot 5, 2 \cdot 3, 2\), which are the invariant factors of the \(\mathbb{Z}\)-module. Which is this \(\mathbb{Z}\)-module?

---

structures: Abelian groups, rings, modules, fields, ordered sets...). For a given type of algebraic structure, the description of all types of isomorphism of the structure is a most important (and hard to attain, in general) objective, called classification. For instance, the theorem of invariant factors yields a classification of finitely generated Abelian groups. The classification of the finite simple groups (having no normal proper subgroups) is one of the great successes of group theory, accomplished in the 1980's.
III.3 Indecomposable finitely generated modules

Exercises

1. Give an example of a ring $R$ and an indecomposable $R$-module that has a decomposable factor module.

2. Prove that $R \cdot M$ is indecomposable $\iff$ the ring $\text{End}_R(M)$ contains no idempotents other than 0 and $1_M$.

3. Give an example of a ring $R$ an indecomposable $R$-module that has decomposable submodules. Can $R$ be a PID? (Hint. Let $K$ be a field, $R = K[X, Y]$, $I = (XY) = XYZ[K, X, Y]$ and $M = R/I$. The idempotents of $\text{End}_R(M)$ are 0 and $1_M$. $XK[X, Y] + YK[X, Y]/I \leq M$ and is decomposable).

4. Let $R$ be a PID, $M$ an $R$-module and $x_1, \ldots, x_n \in M$, with $(o(x_i), o(x_j)) = 1$, $\forall i \neq j$. Then $o(x_1 + \ldots + x_n) = o(x_1)\ldots o(x_n)$. (This generalizes the known fact: if $a, b \in G$, $(G, +)$ an Abelian group, and $(\text{ord } x, \text{ord } y) = 1$, then $\text{ord}(x + y) = \text{ord}(x)\cdot\text{ord}(y)$).

5. Let $R$ be a PID, $M$ an $R$-module with $\text{Ann}_R(M) = R\cdot r$, $r \neq 0$. Then $o(x)| r$, $\forall x \in M$.

6. Let $R$ be a PID and $M$ an $R$-module (not necessarily finitely generated) with $\text{Ann}_R(M) \neq (0)$. We want to prove that $M$ is a direct sum of cyclic submodules. Prove that:

   a) For any $N \leq R \cdot M$, $\text{Ann}_R(M/N) \neq (0)$.

   b) There exists $y \in M$ such that $\text{Ann}_R(y) = \text{Ann}_R(M)$. (Hint. Let $\text{Ann}_R(M) = R\cdot r$, $r = p_1^{a_1}\ldots p_n^{a_n}$, with $p_i$ prime in $R$ and $a_i \in \mathbb{N}^*$. Let $b_i = \max \{ b \in \mathbb{N} | \exists x \in M$ with $p_i^b | o(x) \} (b_i \leq a_i)$. For any $i$, $\exists x_i \in M$ with $o(x_i) = p_i^{b_i}$. Set $y = \sum x_i$. $o(\sum x_i) = \prod o(x_i)$ by ex. 4. Also, $r = \prod p_i^{b_i}$, so $b_i = a_i$)

   c) Let $\mathcal{C} = \{ C \leq M | C$ is a direct sum of cyclic submodules and satisfies condition (*)$\}$, where the condition (*) is:

      $\forall s \in R$, $\forall x \in M$, if $sx \in C$, then $\exists x_0 \in C$ with $sx = sx_0$. (*)

      $\mathcal{C}$ is nonempty, inductively ordered, so it has a maximal element $F$. 
d) \( \forall C \in \mathcal{C}, \text{ if } C \neq M, \text{ then there exists } D \in \mathcal{C} \text{ such that } C \subseteq D. \) (Ind. Apply b) to \( M/C \) and obtain \( y \in M \) such that \( \text{Ann}_R(M/C) = \text{Ann}_R(y) =: \mu. \) So \( \mu y \in C \) and let \( y_0 \in C \) with \( \mu y = \mu y_0, \) given by (*). Then set \( D := C + R(y - y_0). \)

e) \( F = M. \)

### III.4 The endomorphisms of a finite dimensional vector space

_Throughout this section, \( K \) is a field, \( K[X] \) is the polynomial ring in the indeterminate \( X \) with coefficients in \( K, \) \( V \) is a finite dimensional \( K \)-vector space and \( u : V \to V \) is a \( K \)-endomorphism (also called linear transformation or operator)._

The idea of the theory we develop here is the following: _The endomorphism \( u \) defines a natural structure of \( K[X] \)-module on \( V. \) Because \( K[X] \) is a PID, we can apply the invariant factors theorem to obtain information on this \( K[X] \)-module (therefore on the endomorphism \( u \)). Furthermore, keeping in mind the connection between endomorphisms and matrices, the results obtained translate into matrix language._

**4.1 Definition.** Let \( u \in \text{End}_K(V). \) We endow the Abelian group \((V, +)\) with a structure of \( K[X] \)-module (depending on \( u)\):

\[
\forall f = a_0 + a_1X + \ldots + a_nX^n \in K[X], \ \forall v \in V, \text{ define:}
\]

\[
f \cdot v := a_0v + a_1u(v) + \ldots + a_nu^n(v) \in V,
\]

where \( u^n = u \circ \ldots \circ u \) (\( n \) times).

In other words, \( f \cdot v \) is the image of the vector \( v \) by the endomorphism \( f(u) := a_0id + a_1u + \ldots + a_nu^n, \) where \( id \) is the identity automorphism of \( V. \)
Checking the module axioms is simple. It is worth mentioning that the axioms 
\[(fg) \cdot v = (f) \cdot (g \cdot v) \text{ and } (f + g) \cdot v = f \cdot v + g \cdot v \forall f, g \in K[X],
\]
\[\forall v \in V,
\]
are a consequence of the fact that:
\[(fg)(u) = f(u)g(u) \text{ and } (f + g)(u) = f(u) + g(u), \forall f, g \in K[X]
\]
This means that the “evaluation map in \(u\)” from \(K[X]\) to \(\text{End}_K(V)\),
\[f \mapsto f(u), \forall f \in K[X],
\]
is a ring homomorphism (more exactly, a \(K\)-algebra homomorphism).

The external operation defined above extends to \(K[X] \times V\) the external operation (defined on \(K \times V\)) of the \(K\)-vector space \(V\).

Note that the \(K[X]\)-module structure defined on \(V\) depends strongly on the endomorphism \(u\).

Let \(V_u\) denote the \(K[X]\)-module \(V\) defined by the endomorphism \(u\), as above.

4.2 Remark. Let \(u \in \text{End}_K(V)\). The universality property of the polynomial ring \(K[X]\) ensures the existence of a unique \(K\)-algebra homomorphism \(\eta : K[X] \rightarrow \text{End}_K(V)\) such that \(\eta(X) = u\). If \(f = a_0 + a_1X + \ldots + a_nX^n \in K[X]\),
\[\eta(f) = f(u) = \eta(a_0 + a_1X + \ldots + a_nX^n) = a_0 \text{id}_V + a_1u + \ldots + a_nu^n.
\]
The homomorphism \(\eta\) defines a \(K[X]\)-module structure on \(V\) (see remark II.1.4) that is the same as the one defined above.

Can distinct endomorphisms of \(V\) define “the same” \(K[X]\)-module structure on \(V\)?

4.3 Proposition. Let \(u, w \in \text{End}_K(V)\). Then the \(K[X]\)-modules \(V_u\) and \(V_w\) are isomorphic if and only if there exists a \(K\)-automorphism \(\phi\) of \(V\) such that \(\phi \cdot u \cdot \phi^{-1} = w\).

Proof. Let \(\phi : V_u \rightarrow V_w\) be a \(K[X]\)-isomorphism. Let \(\cdot_u\) denote the external operation of the \(K[X]\)-module \(V_u\). For any \(v \in V_u\), \(X \cdot_u v = u(v)\) and, \(\forall v \in V_w\), \(X \cdot_w v = w(v)\). We have \(\phi(X \cdot u v) = X \cdot w \phi(v)\), so \(\phi(u(v)) = w(\phi(v))\), i.e. \(\phi \cdot u = w \cdot \phi\). It is clear that \(\phi\) is also a \(K\)-automorphism.
III. Finitely generated modules over principal ideal domains

If \( \varphi : V \rightarrow V \) is a \( K \)-isomorphism with \( \varphi \circ u = w \circ \varphi \), then \( \varphi \) is a \( K[X] \)-isomorphism from \( V_u \) to \( V_w \). Indeed, \( \varphi \circ u = w \circ \varphi \) is written \( \varphi(X^n \cdot u v) = X^n \cdot w \varphi(v) \), \( \forall n \in \mathbb{N} \); since \( \varphi \) is \( K \)-linear,

\[
\varphi((a_0 + a_1X + \ldots + a_nX^n) \cdot u v) = a_0 \cdot w \varphi(v) + a_1X \cdot w \varphi(v) + \ldots + a_nX^n \cdot w \varphi(v),
\]

for any \( a_0 + a_1X + \ldots + a_nX \in K[X] \).

4.4 Definition. a) The endomorphisms \( u, w \in \text{End}_K(V) \) are called similar if there exists \( \varphi \in \text{Aut}_K(V) \) such that \( \varphi \circ u \circ \varphi^{-1} = w \). Notation: \( u \approx w \).

b) Let \( n \in \mathbb{N} \) and \( A, B \in M_n(K) \). We call the matrices \( A \) and \( B \) similar (denoted by \( A \approx B \)) if there exists an invertible matrix \( U \in M_n(K) \) such that \( B = UAU^{-1} \).

4.5 Remark. Let \( _K V \) be finite dimensional and let \( v = (v_1, \ldots, v_n) \) be a basis of \( V \). The endomorphisms \( u, w \in \text{End}_K(V) \) are similar if and only if the matrices \( M_v(u) \) and \( M_v(w) \) are similar. To see that, recall that the rings \( \text{End}_K(V) \) and \( M_n(K) \) are (anti-)isomorphic by the mapping \( u \mapsto M_v(u) \).

A typical example of similar matrices is given by the matrices of an endomorphism in various bases of \( V \).

The similarity relation on \( \text{End}_K(V) \) and the similarity relation on \( M_n(K) \) are equivalence relations.

4.6 Definition. Let \( u \in \text{End}_K(V) \) and \( W \leq _K V \). The \( K \)-subspace \( W \) of \( V \) is called invariant relative to \( u \) (or \( u \)-invariant) if \( u(W) \subseteq W \). We say that \( V \) is indecomposable relative to \( u \) (or \( u \)-indecomposable) if \( V \) cannot be written as a direct sum of proper \( u \)-invariant subspaces.

4.7 Proposition. Let \( W \) be a nonempty set of \( V \). Then: \( W \) is an \( u \)-invariant subspace \( \iff \forall w \in W, X \cdot w \in W \iff W \) is a \( K[X] \)-submodule of \( V_u \).
III.4 The endomorphisms of a finite dimensional vector space

Proof. \( u(W) \subseteq W \) if and only if \( \forall w \in W, u(w) \in W \Leftrightarrow Xw \in W \). By induction, \( X^n w \in W, \forall n \in \mathbb{N} \); since \( W \) is a linear subspace in \( V \), this implies \( (a_0 + a_1 X + \ldots + a_n X^n)w \in W, \forall a_i \in K \). The converse is an easy exercise. \( \square \)

The following statements (whose proofs are left to the reader) are further examples of translations from vector space language to \( K[X] \)-module language.

“\( V \) is the direct sum of the \( u \)-invariant subspaces \( V_1, \ldots, V_m \)” \( \Leftrightarrow \) “\( K[X]V_u \) is the direct sum of the \( K[X] \)-submodules \( V_1, \ldots, V_m \)”.

“There exists \( v \in V \) such that \( V \) is generated by \( \{u^i(v) | i \in \mathbb{N}\} \)” \( \Leftrightarrow \) “\( V_u \) is a cyclic \( K[X] \)-module (generated by \( v \)).” (In this case, \( u \) is called a cyclic endomorphism).

“\( V \) is \( u \)-indecomposable” \( \Leftrightarrow \) “\( V_u \) is an indecomposable \( K[X] \)-module”.

4.8 Proposition. a) The \( K[X] \)-module \( V_u \) is a finitely generated torsion module.

b) \( V \) is \( u \)-indecomposable if and only if there exists \( v \in V_u \) such that \( V_u = K[X]v \), with \( o(v) = p^k \), where \( p \) is an irreducible polynomial in \( K[X] \) and \( k \in \mathbb{N}^* \).

Proof. a) Any system of generators \( \{v, u(v), \ldots, u^n(v)\} \) of \( V \) is also a system of generating elements for \( V_u \). If \( \dim_K V = n \in \mathbb{N}^* \), then, \( \forall v \in V \), the vectors \( v, u(v), \ldots, u^n(v) \) cannot be linearly independent. Consequently, there exist \( a_0, a_1, \ldots, a_n \in K \), not all zero, such that \( a_0 v + a_1 u(v) + \ldots + a_n u^n(v) = 0 \Leftrightarrow (a_0 + a_1 X + \ldots + a_n X^n) \cdot v = 0 \).

b) \( V \) is \( u \)-indecomposable if and only if \( V_u \) is an indecomposable \( K[X] \)-module. But \( V_u \) is a torsion, finitely generated \( K[X] \)-module. Apply now Prop. III.3.7. \( \square \)

We apply to the \( K[X] \)-module \( V_u \) the structure theorems from III.2 and III.3:
4.9 Proposition. If \( \dim_K V = n \in \mathbb{N}^* \), then there exists \( m \in \mathbb{N}^* \) and \( v_1, \ldots, v_m \in V_u \) such that \( V_u \) is a direct sum of \( u \)-invariant subspaces
\[
V_u = K[X]v_1 \oplus \ldots \oplus K[X]v_m, \text{ cu } o(v_1) | \ldots | o(v_m).
\]

The natural number \( m \) and the monic polynomials \( o(v_1), \ldots, o(v_m) \in K[X] \) with the properties above are uniquely determined (\( o(v_1), \ldots, o(v_m) \) are the invariant factors of the \( K[X] \)-module \( V_u \)).

\( V_u \) is also a direct sum of indecomposable submodules:
\[
V_u = K[X]w_1 \oplus \ldots \oplus K[X]w_t,
\]
where \( w_i \in V_u \) and \( o(w_i) \) are powers of irreducible polynomials in \( K[X] \), \( 1 \leq i \leq t \). The natural number \( t \) and the monic polynomials \( o(w_1), \ldots, o(w_t) \in K[X] \) are uniquely determined (\( o(w_1), \ldots, o(w_t) \) are the elementary divisors of the \( K[X] \)-module \( V_u \)).

4.10 Definition. If \( \dim_K V = n \) and \( u \in \text{End}_K(V) \), the monic polynomial that generates \( \text{Ann}_{K[X]}(V_u) \) is called the minimal polynomial of the endomorphism \( u \) (and is denoted by \( \mu_u \)). The monic polynomials that are the invariant factors (respectively the elementary divisors) of the \( K[X] \)-module \( V_u \) are called the invariant factors (respectively the elementary divisors) of the endomorphism \( u \). The same terminology is used for matrices: choosing a basis \( v \) in \( V \), the invariant factors of a matrix \( A \in M_n(K) \) are the invariant factors of the unique endomorphism \( u \) with \( M_v(u) = A \) (similarly for the minimal polynomial, the elementary divisors).

4.11 Proposition. Two endomorphisms (matrices) are similar \( \iff \) they have the same invariant factors \( \iff \) they have the same elementary divisors.

Proof. Let \( u, w \in \text{End}_K(V) \). We have: \( M_v(u) \cong M_v(w) \iff u \cong w \iff V_u \cong_{K[X]} V_w \) (by 4.3) \( \iff V_u \) and \( V_w \) have the same invariant factors \( \iff V_u \) and \( V_w \) have the same elementary divisors.
4.12 Remark. With the notations in 4.9, \( \mu_u \) is the highest degree invariant factor of \( u \). If \( f \in K[X] \) is monic, the following properties are equivalent:

a) \( f = \mu_u \).

b) \( f(u) = 0 \) and \( \forall g \in K[X], g(u) = 0 \) implies \( f \mid g \).

c) \( f(u) = 0 \) and \( \forall g \in K[X], g \neq 0, g(u) = 0 \), implies \( \deg f \leq \deg g \).

The proof is easy, using the definitions. Note that, unlike the minimal polynomial of an algebraic element in a field extension, the minimal polynomial of an endomorphism is not necessarily irreducible.

The next result translates in matrix language the fact that \( V \) is a direct sum of \( u \)-invariant subspaces.

4.13 Proposition. a) Let \( V = V_1 \oplus \ldots \oplus V_m \), where \( V_1, \ldots, V_m \) are \( u \)-invariant subspaces. If \( v_i \) is a basis in \( V_i \), \( 1 \leq i \leq m \), then \( v_1 \cup \ldots \cup v_m =: v \) is a basis\(^7\) in \( V \). If \( A_i \) is the matrix of the restriction of \( u \) to \( V_i \), in the basis \( v_i, 1 \leq i \leq m \), then the matrix of \( u \) in the basis \( v \) is (written on blocks):

\[
M_v(u) = \begin{bmatrix}
A_1 & 0 \\
\vdots & \ddots \\
0 & A_m
\end{bmatrix},
\]

b) Conversely, if the matrix of \( u \) in a basis \( v \) is of the form above, then the rows of the block \( A_i \) correspond to a set of vectors in \( v \) that generate an \( u \)-invariant subspace \( V_i \) (\( 1 \leq i \leq m \)) and \( V = V_1 \oplus \ldots \oplus V_m \).

Proof. a) It is clear that \( v \) is a basis in \( V \) (see also II.4.11). To keep notations manageable, suppose \( m = 2 \) and \( v_1 = (e_1, \ldots, e_p) \), \( v_2 = (f_1, \ldots, f_q) \), \( p + q = n = \dim V \). Then \( v = (e_1, \ldots, e_p, f_1, \ldots, f_q) \). Since \( V_1, V_2 \) are \( u \)-invariant, \( u(e_i) \) is a linear combination of \( e_1, \ldots, e_p \), and \( u(f_j) \) is a linear combination of \( f_1, \ldots, f_q \). Writing the matrix of \( u \) in the basis \( v \),

\(^7\) We totally order the vectors in the basis \( v \), by sequencing the elements of the bases \( v_1, \ldots, v_m \), in this order.
III. Finitely generated modules over principal ideal domains

\[
M_v(u) = \begin{bmatrix}
A_1 & 0 \\
0 & A_2
\end{bmatrix}.
\]

b) The task of detailing the proof is left to the reader.

We want to find a basis \( v \) of \( V \) such that \( M_v(u) \) has as “simple” a form as possible. Since \( V_u \) is a direct sum of indecomposable submodules (theorem 4.9) the previous result allows us to study the restriction of \( u \) to each of the \( u \)-invariant subspaces in the direct sum. It is thus natural to study first the case in which \( V_u \) is indecomposable: \( V_u = K[X]v \), for some \( v \in V \), \( o(v) = p^k \), \( p \) irreducible in \( K[X] \), \( k \in \mathbb{N}^* \).

4.14 Definition. If \( p \in K[X] \), \( p = X^r - a_{r-1}X^{r-1} - \ldots - a_1X - a_0 \), define the \( r \times r \) matrices with entries in \( K \):

\[
C_p = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
a_0 & a_1 & a_2 & \cdots & a_{r-1}
\end{bmatrix}, \quad N = \begin{bmatrix}
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots & 0
\end{bmatrix}
\]

The matrix \( C_p \) is called the matrix companion of the polynomial \( p \). Define the \( r \times r \) matrix (written in block form):

\[
J(p^k) = \begin{bmatrix}
C_p & N & 0 & \cdots & 0 \\
0 & C_p & N & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & C_p \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\end{bmatrix} \in M_{rk}(K)
\]

\( J(p^k) \) is called the Jordan cell \(^8\) associated to the polynomial \( p^k \).

---

\(^8\) Camille Jordan (1838-1922), French mathematician.
A matrix (written in block form) having on the diagonal Jordan cells (and 0 elsewhere), i.e. a matrix of the form

\[
\begin{bmatrix}
J(p_{1}^{k_{1}}) & 0 \\
J(p_{2}^{k_{2}}) & \\
\ddots & \\
0 & J(p_{t}^{k_{t}})
\end{bmatrix},
\]

where \(p_{1}, \ldots, p_{t}\) are monic irreducible polynomials in \(K[X]\), is called a Jordan canonical matrix\(^9\) over \(K\).

4.15 Proposition. a) Suppose \(V_{u}\) is a indecomposable \(K[X]-\)module and \(v \in V_{u}\) is such that \(V_{u} = K[X]v\), with \(o(v) = \mu_{u} = p^{k}\), where \(p = X^{r} - a_{r-1}X^{r-1} - \ldots - a_{1}X - a_{0} \in K[X]\) is irreducible of degree \(r\) and \(k \in \mathbb{N}^{*}\). Then \(\dim_{K}V = rk\) and there exists a basis \(v\) of \(KV\) such that the matrix of \(u\) in \(v\) is the Jordan cell \(J(p^{k})\) associated to \(p^{k}\).

b) In the general case, let \(p_{1}^{k_{1}}, \ldots, p_{t}^{k_{t}}\) be the elementary divisors of \(u\), with \(p_{1}, \ldots, p_{t}\) irreducible and monic in \(K[X]\). Then there exists a basis of \(V\) in which the matrix of \(u\) is the Jordan canonical matrix:

\[
J = \begin{bmatrix}
J(p_{1}^{k_{1}}) & 0 \\
J(p_{2}^{k_{2}}) & \\
\ddots & \\
0 & J(p_{t}^{k_{t}})
\end{bmatrix}.
\]

Proof. a) We exhibit a basis \(e = (e_{0}, \ldots, e_{kr-1})\) such that \(M_{e}(u) = J(p^{k})\):

---

\(^9\) In other texts this matrix is called a rational canonical matrix, the name Jordan canonical matrix being given only if \(p_{i}\) are polynomials of degree 1.
\[ e_0 = v; \quad e_1 = X_1 \cdot v = u(e_0); \ldots \quad e_{r-1} = X_{r-1} \cdot v = u(e_{r-2}); \]
\[ e_r = p \cdot v; \quad e_{r+1} = Xp \cdot v = u(e_r); \ldots \quad e_{2r-1} = X^{r-1} p \cdot v = u(e_{2r-2}); \]
\[ \ldots \]
\[ e_{(k-1)r} = p^{k-1} \cdot v; \quad e_{(k-1)r+1} = Xp \cdot v = u(e_{(k-1)r+1}); \ldots ; e_{kr-1} = X^{r-1} p^{k-1} \cdot v = u(e_{kr-2}). \]

The next lemma proves that \( e = (e_0, \ldots, e_{kr-1}) \) is a basis:

**Lemma.** If \( u \in \text{End}_K(V) \) is an endomorphism such that \( V_u = K[X]v \) for some \( v \in V \), and \( f = o(v) \), \( \deg f = n \), then, for any \( g_0, \ldots, g_{n-1} \in K[X] \), with \( \deg g_i = i, 1 \leq i \leq n \), the vectors
\[ g_0 \cdot v, \ldots, g_{n-1} \cdot v \]
form a basis of \( V \).

**Proof of the lemma.** \( V_u \cong K[X]/(f) \) (\( K[X] \)-module isomorphism, so also a \( K \)-vector space isomorphism), hence \( \dim K V = \dim K K[X]/(f) = \deg f = n \). The vectors \( g_0 \cdot v, \ldots, g_{n-1} \cdot v \) are linearly independent: if
\[ a_0 g_0 \cdot v + \ldots + a_{n-1} g_{n-1} \cdot v = 0, \]
with \( a_i \in K \), then \( h \cdot v = 0 \), where \( h = a_0 g_0 + \ldots + a_{n-1} g_{n-1} \). Since \( o(v) = f \), and \( \deg h < n \), we have \( f|/h \), hence \( h = 0 \).

But the polynomials \( g_0, \ldots, g_{n-1} \) are linearly independent in the \( K \)-vector space \( K[X] \), being of distinct degrees, so \( a_0 = \ldots = a_{n-1} = 0 \).

The \( n \) elements \( g_0 \cdot v, \ldots, g_{n-1} \cdot v \) are thus linearly independent in \( V \), whose dimension is \( n \), which means they are a basis.

We get back to proving that \( \text{M}_e(u) = J(p^k) \). If \( 1 \leq i < k \), we have:
\[ u(e_{ir-1}) = X_i X_{r-1} \cdot p^{i-1} \cdot v = X_r p^{i-1} \cdot v = (p + a_0 + a_1 X + \ldots + a_{r-1} X^{r-1}) p^{i-1} \cdot v = p^i \cdot v + a_0 p^{i-1} \cdot v + a_1 X p^{i-1} \cdot v + \ldots + a_{r-1} X^{r-1} p^{i-1} \cdot v = e_{ir} + a_0 e_{(i-1) r} + a_1 e_{(i-1) r+1} + \ldots + a_{(i-1) r} \]

If \( i = k \),
\[ u(e_{kr-1}) = a_0 e_{(k-1) r} + a_1 e_{(k-1) r+1} + \ldots + a_{(k-1) r} e_{(k-1) r+r-1} \]

since \( p^k \cdot v = 0 \).

These equalities, together with the relations (0), \( \ldots, (k - 1) \), prove the claim.

b) Decompose \( V_u \) as a direct sum of \( u \)-indecomposable \( u \)-invariant subspaces (see 4.9). By a), each such subspace has a basis in which the restriction of \( u \) has the matrix of the form \( J(p^k) \), with \( p^k \) elementary divisor of \( u \). Apply now Prop. 4.13.
4.16 Corollary. Any matrix \( A \in M_n(K) \) is similar to a Jordan canonical matrix. If the elementary divisors of \( A \) are \( p_1^{k_1}, \ldots, p_t^{k_t} \), then \( A \approx J \), where \( J \) is the matrix at 4.15.b).

How do we find the elementary divisors of an endomorphism \( u \) (of a matrix \( A \))? This amounts to finding its invariant factors. The next theorem says: take the matrix \( XI - A \) and find its Smith normal form. The non constant polynomials on the diagonal are then the invariant factors of \( A \).

4.17 Theorem. Let \( v = (v_1, \ldots, v_n) \) be a basis in \( V \) such that \( M_v(u) =: A = (a_{ij}) \in M_n(K) \). Then the invariant factors of \( u \) (of the matrix \( A \)) are the polynomials of degree \( > 0 \) on the diagonal of the diagonally canonic matrix \( D \in M_n(K[X]) \), arithmetically equivalent to the matrix

\[
XI - A = \begin{bmatrix}
X - a_{11} & -a_{12} & \cdots & -a_{1n} \\
-\ a_{21} & X - a_{22} & \cdots & -a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
-\ a_{n1} & \ -a_{n2} & \cdots & X - a_{nn}
\end{bmatrix} \in M_n(K[X]).
\]

Proof. Recall the proof of the invariant factors theorem (2.3): since \((v_1, \ldots, v_n)\) generates \( K[X]V_u \), take a \( K[X] \)-module \( E \), free of basis \( e = (e_1, \ldots, e_n) \) and the \( K[X] \)-homomorphism \( \phi : E \to V_u \), \( \phi(e_i) = v_i \), \( 1 \leq i \leq n \). The homomorphism \( \phi \) is surjective, \( \ker \phi =: F \) is a free submodule in \( E \) and \( E/F \cong V_u \). Since \( V_u \) is torsion, rank \( F = n \) and there exist two bases \( e = (e_1, \ldots, e_n) \) in \( E \) and \( \phi = (\phi_1, \ldots, \phi_n) \) in \( F \), such that \( \phi_i = d_i e_i \), \( d_i \in K[X] \), \( d_1|d_2| \ldots |d_n \). The invariant factors of \( V_u \) are the noninvertible polynomials among \( d_1, d_2, \ldots, d_n \); (see the proof of 2.3)

How do we find \( d_1, d_2, \ldots, d_n \)? For any basis \( f \) in \( F \), there exists a matrix \( B \in M_n(K[X]) \) such that \( f = Be \). Then the diagonally canonical matrix \( D \in M_n(K[X]) \), arithmetically equivalent to \( B \), is exactly \( D = \text{diag} (d_1, d_2, \ldots, d_n) \). Thus, we need such a matrix \( B \). Let

\[
f_i = : Xe_i - (a_{i1}e_1 + a_{i2}e_2 + \ldots + a_{in}e_n) \in E, \ 1 \leq i \leq n.
\]
If \( f = (f_1, f_2, \ldots, f_n) \), then these relations can be written \( f = (XI - A)e \).

**Lemma.** \( f =: (f_1, f_2, \ldots, f_n) \) is a basis in \( F = \text{Ker } \varphi \).

*Proof of the lemma.* For any \( 1 \leq i \leq n \),
\[
u(i) = a_{i1}v_1 + a_{i2}v_2 + \ldots + a_{in}v_n
\]

Thus,
\[
\varphi(f_i) = Xv_i - (a_{i1}v_1 + a_{i2}v_2 + \ldots + a_{in}v_n)
= u(v_i) - (a_{i1}v_1 + a_{i2}v_2 + \ldots + a_{in}v_n) = 0.
\]

This shows that \( f_i \in \text{Ker } \varphi = F \). Let us prove that \( f \) is a system of generators for \( F \). Note that:
\[
Xe_i = f_i + a_{i1}e_1 + a_{i2}e_2 + \ldots + a_{in}e_n, \quad 1 \leq i \leq n.
\]

So, \( X^2e_i = Xf_i + a_{i1}Xe_1 + a_{i2}Xe_2 + \ldots + a_{in}Xe_n \). Using (*), we obtain:
\[
X^2e_i = \sum q_jf_j + \sum r_je_j, \quad \text{for some } q_j \in K[X], \ r_j \in K, \ 1 \leq j \leq n.
\]

By induction, one easily sees that, \( \forall m \in \mathbb{N}^* \), \( X^me_i \) is expressed as:
\[
X^me_i = \sum q_jf_j + \sum r_je_j, \quad \text{for some } q_j \in K[X], \ r_j \in K, \ 1 \leq j \leq n. \quad (**)
\]

From (**), we deduce that, \( \forall g \in K[X], \ ge_i \) can be written in the same form. So, if \( y = \sum g_ie_i \in F \), with \( g_i \in K[X], \ 1 \leq i \leq n \), then
\[
y = \sum g_ie_i = \sum q_if_i + r,
\]
for some \( q_j \in K[X] \), and \( r \in E \) is of the form \( \sum c_ie_i \), with \( c_i \in K \). But \( r = y - \sum q_if_i \in F \), so \( \varphi(r) = \varphi(\sum c_ie_i) = \sum c_ie_i = 0 \). Since \( v \) is a basis in \( V, \ c_i = 0, \ 1 \leq i \leq n \), hence \( r = 0 \). Finally, we obtain \( y = \sum q_if_i \).

Let us prove the linear independence. Suppose \( \sum g_if_i = 0 \), with \( g_i \in K[X], \ 1 \leq i \leq n \). Using (*), it follows that
\[
\sum g_iXe_i = \sum g_i(\sum a_{ij}e_j) = \sum (\sum a_{ij}g_i)e_j.
\]

Since \( (e_1, \ldots, e_n) \) is a basis,
\[
g_1X = \sum a_{ij}g_j, \quad 1 \leq i \leq n.
\]

Let \( g_1 \) be the polynomial of maximal degree among the \( g_i \) (relabel if necessary). So, \( \deg g_i \leq \deg g_1, \ 1 \leq i \leq n \).

If \( g_1 \neq 0 \), \( g_1X = \sum a_{j1}g_j \) implies
\[
\deg(\sum a_{j1}g_j) \leq \max_j(\deg(a_{j1}g_j)) \leq \max_j(\deg g_j) < 1 + \deg g_1 = \deg g_1X,
\]
contradiction. This shows that \( (f_1, f_2, \ldots, f_n) \) is linearly independent.
We continue the proof of the theorem. The relations

\[ f_i = Xe_i - (a_{i1}e_1 + a_{i2}e_2 + \ldots + a_{im}e_m) \]

show that \( f = (XI - A)e \). If \( XI - A \sim D = \text{diag}(d_1, d_2, \ldots, d_n) \), with \( d_1|d_2|\ldots|d_n \), then the invariant factors of \( V_u \) (of the endomorphism \( u \)) are \( d_k, d_{k+1}, \ldots, d_n \), where \( k = \min \{ i \mid d_i \text{ noninvertible} \} \).

4.18 Definition. Let \( A \in M_n(K) \). The polynomial \( f_A := \det(XI - A) \in K[X] \) is called the characteristic polynomial of the matrix \( A \). If \( u \) is an endomorphism of the \( K \)-vector space \( V \) and \( A \) is the matrix of \( u \) (in some basis), then the characteristic polynomial \( f_u \) of \( u \) is by definition \( f_A \). This definition is correct: Two similar matrices \( A \) and \( B \) have the same characteristic polynomial: if \( B = SAS^{-1} \), for some \( S \in GL_n(K) \), then:

\[
  f_B = \det(XI - SAS^{-1}) = \det(S(XI - A)S^{-1}) = \\
  = \det(S) \cdot \det(XI - A) \cdot \det(S^{-1}) = f_A,
\]

We used the fact that the matrix \( XI \) commutes with any matrix in \( M_n(K[X]) \).

4.19 Remark. Let \( A = (a_{ij}) \in M_n(K) \) and let

\[
  f_A = \det(XI - A) = X^n - c_1 X^{n-1} + c_2 X^{n-2} + \ldots + (-1)^n c_n,
\]

with \( c_1, \ldots, c_n \in K \).

Writing the definition of \( \det(XI - A) \) and arranging the terms by the like powers of \( X \), the coefficients \( c_1, \ldots, c_n \) are:

\[
  c_1 = a_{11} + a_{22} + \ldots + a_{nn} =: \text{Tr}(A) \text{ (called the trace of } A),
\]

\[
  c_n = \det(A).
\]

More generally, \( c_k \) (\( 1 \leq k \leq n \)) is the sum of the minors of order \( k \) of \( A \) on the main diagonal (i.e., the minors obtained by selecting \( k \) rows \( \{i_1, \ldots, i_k\} \) and the columns with the same indices \( \{i_1, \ldots, i_k\} \) of the matrix \( A \)). There are \( \binom{n}{k} \) such minors, one for each choice of a subset of \( k \) indices from \( \{1, 2, \ldots, n\} \).
The same terminology applies for an endomorphism \( u \in \text{End}_K(V) \), \( \dim V = n \), whose matrix is \( A \) (in some basis of \( V \)). Its characteristic polynomial is \( f_u = f_A \), and the coefficients \( c_1, \ldots, c_n \) are uniquely defined by \( u \), as above. \( c_1 =: \text{Tr}(u) = \text{Tr}(A) \) is called the \textit{trace} of \( u \) and \( c_n =: \det(u) = \det(A) \) is called the \textit{determinant} of \( u \).

\[4.20 \text{ Proposition.} \] Let \( A, B \in M_n(K) \). Then \( A \) and \( B \) are similar matrices if and only if \( XI - A \) and \( XI - B \) are arithmetically equivalent matrices in \( M_n(K[X]) \).

\textbf{Proof.} If \( A \approx B \), then there exists \( S \in GL(n, K) \) such that \( B = S^{-1}AS \). Then \( XI - B = S^{-1}(XI - A)S \), so \( XI - A \sim XI - B \), since obviously \( S \in GL(n, K[X]) \).

Suppose now that \( XI - A \sim XI - B \) and let \( D \in M_n(K[X]) \) be the canonical matrix with \( D \sim (XI - A) \sim (XI - B) \). So, \( A \) and \( B \) have the same invariant factors: the polynomials of degree \( > 0 \) on the diagonal of \( D \), according to 4.17. Thus, \( A \) and \( B \) have the same elementary divisors. By 4.15, \( A \) and \( B \) are similar with the same Jordan canonical matrix. \( \square \)

Let \( n \in \mathbb{N}^* \). The following results are about matrices in \( M_n(K) \), but they can be translated in statements on endomorphisms of a \( K \)-vector space of dimension \( n \).

\[4.21 \text{ Proposition.} \] The characteristic polynomial of a matrix \( A \) is the product of the invariant factors of \( A \) (and it is equal to the product of the elementary divisors of \( A \)).

\textbf{Proof.} The matrix \( XI - A \) is arithmetically equivalent to the matrix in Smith normal form \( D = \text{diag}(1, \ldots, 1, d_1, \ldots, d_m) \in M_n(K[X]) \), where \( d_1, \ldots, d_m \) are the invariant factors of \( A \). There exist \( S, T \in U(M_n(K[X])) \) (i.e. \( \det S, \det T \in K^* \)) such that \( XI - A = SDT \). We have

\[ d_1 \cdot \ldots \cdot d_m = \det D = \det(S(XI - A)T) = \det S f_A \det T \]
This means that the monic polynomials $d_1 \cdots d_m$ and $f_A$ differ by the factor $\det S \det T \in K^*$, which shows that they are equal. On the other hand, it is clear that the product of the elementary divisors equals the product of the invariant factors.

4.22 Corollary. a) (The Cayley-Hamilton theorem)\textsuperscript{10} Any matrix $A \in M_n(K)$ is a root of its characteristic polynomial: $f_A(A) = 0$.

b) (The Frobenius theorem)\textsuperscript{11} The characteristic polynomial and the minimal polynomial of a matrix $A \in M_n(K)$ have the same irreducible factors in $K[X]$.

Proof. Let $d_1, \ldots, d_m \in K[X]$ be the invariant factors of $A$.

a) The minimal polynomial of $A$ is $d_m$ (the invariant factor of highest degree). So, $d_m(A) = 0$ and $d_m | f_A$, thus $f_A(A) = 0$.

b) This is a consequence of $d_1 \cdots d_m = f_A$ and $d_1 \cdots | d_m$. □

Given a matrix $A$, is the Jordan canonical matrix that is similar to $A$ uniquely determined? Of course, by reordering the elementary divisors, diverse Jordan canonical matrices are obtained. The next proposition shows that these are all the Jordan canonical matrices similar to $A$.

4.23 Proposition. a) Let $p \in K[X]$ be monic and irreducible and let $k \in \mathbb{N}^*$. The Jordan cell $J(p^k)$ has only one elementary divisor, namely $p^k$.

b) Let $A$ be a Jordan canonical matrix whose diagonal is made up by the Jordan cells $J(p_i^k)$, where $p_i \in K[X]$ are monic and irreducible, $1 \leq i \leq t$. Then the elementary divisors of $A$ are $p_i^k$, $1 \leq i \leq t$.

\textsuperscript{10} Arthur Cayley (1821-1895), Sir William Rowan Hamilton (1805-1865), British mathematicians.

\textsuperscript{11} Ferdinand Georg Frobenius (1849-1917), German mathematician.
c) Let $A, B \in M_n(K)$ be Jordan canonical matrices such that $A \approx B$. Then $A$ and $B$ have the same Jordan cells, perhaps in different order.

**Proof.** a) Let $V := K[X]/(p^k)$. $V$ is a $K[X]$-module (being a factor module of $K[X]$) and a $K$-vector space of dimension $k \cdot \deg p =: n$. Let $u \in \text{End}_K(V)$, $u(y) = X \cdot y$, $\forall y \in V$ (the dot denotes the external $K[X]$-module operation of $V$). The endomorphism $u$ defines on $V$ a structure of a $K[X]$-module, $Vu$, as in definition 4.1. It is easy to see that these two structures of $K[X]$-module coincide and that $V = K[X]v$, where $v := 1 + (p^k) \in V$. So, $V = Vu$ is an indecomposable $K[X]$-module, and its only elementary divisor is $o(v) = p^k$. There exists a basis (as in 4.15) in which $u$ has the matrix $J(p^k)$. So $J(p^k)$ has the same elementary divisors as $Vu$, namely only $p^k$.

b) There exists $Kv$ and $u \in \text{End}_K(V)$, whose matrix is $A$ (in some basis). Then $V$ is written as a direct sum of $u$-invariant subspaces: $Vu = V_1 \oplus \ldots \oplus V_t$, $V_i$ being the $u$-invariant subspace corresponding to the Jordan cell $J(p^k_i)$. Let $u_i$ be the restriction of $u$ to $V_i$, $1 \leq i \leq t$. The matrix of $u_i$ is exactly $J(p^k_i)$ and the first part of the proof shows that $p^k_i$ is the only elementary divisor of $u_i$. The elementary divisors of $u$ are obtained writing down all the elementary divisors of the restrictions $u_i$, $1 \leq i \leq t$, namely $p^k_i$, $1 \leq i \leq t$.

c) If $A \approx B$, then $A$ and $B$ have the same elementary divisors (4.11).

By b), $A$ and $B$ have the same Jordan cells. 

**4.24 Definition.** If $A \in M_n(K)$ and $J$ is a Jordan canonical matrix such that $J \approx A$, then $J$ is called the Jordan canonical form of $A$. The Jordan canonical form of $A$ is uniquely determined up to an order of the Jordan cells on the diagonal.

The classical notions of eigenvector and eigenvalue of an endomorphism are closely connected to its invariant subspaces of dimension one.
4.25 **Definition.** If \( u \in \text{End}_K(V) \), an element \( \lambda \in K \) is called an eigenvalue of \( u \) if there exists a vector \( v \in V, v \neq 0 \), such that
\[
u(v) = \lambda v.\]

Each such vector \( v \) is called an eigenvector of \( u \) for the eigenvalue \( \lambda \).

4.26 **Proposition.** Let \( v \in V, u \in \text{End}_K(V) \) and \( \lambda \in K \). The following statements are equivalent:

a) \( v \) is an eigenvector of \( u \) for the eigenvalue \( \lambda \).

b) Considering \( v \in K[X]V_u, o(v) = X - \lambda \in K[X] \).

c) \( \dim_K(K[X]v) = 1 \) (the submodule of \( V_u \) generated by \( v \) has \( K \)-dimension 1).

**Proof.**

a) \( \Rightarrow b) \) We have \( u(v) = \lambda v \). In \( K[X] \)-module language for \( V_u \), this means \( X \cdot v = \lambda v \), thus \( (X - \lambda) \cdot v = 0 \). So, \( o(v) \mid X - \lambda \), which is irreducible, hence \( o(v) = X - \lambda \). \( (o(v) = 1 \) is impossible, since it implies \( v = 0 \)).

b) \( \Rightarrow a) \) \( o(v) = X - \lambda \) implies \( u(v) = \lambda v \).

b) \( \iff c) \) This follows from \( \dim_K K[X]v = \deg o(v) \). \( \square \)

4.27 **Proposition.** \( \lambda \in K \) is an eigenvalue of \( u \) if and only if \( \lambda \) is a root of \( f_u \), the characteristic polynomial of \( u \).

**Proof.** Suppose \( v \in V \) is an eigenvector of \( u \) for the eigenvalue \( \lambda \).

Then \( o(v) = X - \lambda \); since the minimal polynomial of \( u \), \( \mu_u \), is in \( \text{Ann}_{K[X]}(v) = (X - \lambda) \), we have \( X - \lambda \mid \mu_u \). Also \( \mu_u \mid f_u \), so \( X - \lambda \mid f_u \), which means that \( \lambda \) is a root of \( f_u \). Conversely, if \( f_u(\lambda) = 0 \), then \( X - \lambda \mid f_u \).

Because \( f_u \) and \( \mu_u \) have the same irreducible factors, \( X - \lambda \mid \mu_u \), so \( \mu_u = (X - \lambda)g \) for some \( g \in K[X] \). By 4.9, there exists \( v \in V \) such that \( o(v) = \mu_u \). Then \( o(g \cdot v) = X - \lambda \), i.e. \( g \cdot v \in V \) is an eigenvector for the eigenvalue \( \lambda \).

The reader is invited to give an alternate proof, using facts from the theory of systems linear equations.
Note that if $f_u$ has no roots in $K$, then $u$ has no eigenvalues and no eigenvectors.

We describe now the Jordan cells in the important cases $K = \mathbb{C}$ and $K = \mathbb{R}$.

If $K$ is an algebraically closed field (in particular, $K = \mathbb{C}$), then the irreducible monic polynomials in $K[X]$ are of the form $X - a, a \in K$. Thus the Jordan cell $J((X - a)^k)$ is:

$$J((X - a)^k) = \begin{bmatrix}
a & 1 & 0 & \ldots & 0 & 0 \\
0 & a & 1 & \ldots & 0 & 0 \\
& & & & & \ddots \\
& & & & & \\
0 & 0 & 0 & \ldots & a & 1 \\
0 & 0 & 0 & \ldots & 0 & a
\end{bmatrix} \in M_k(K)$$

If $K = \mathbb{R}$, the monic irreducible polynomials in $\mathbb{R}[X]$ are $X - a, a \in \mathbb{R}$ (and the Jordan cell $J((X - a)^k)$ is the one above), or $X^2 - bX - c$, with $b, c \in \mathbb{R}$ and $b^2 + 4c < 0$, in which case the Jordan cell $J((X^2 - bX - c)^k)$ is:

$$J((X^2 - bX - c)^k) = \begin{bmatrix}
0 & 1 & 0 & 0 \\
c & b & 1 & 0 \\
& 0 & 1 & 0 & 0 \\
& c & b & 1 & 0 \\
& & & & \ddots \\
& & & & & \ddots \\
0 & 0 & 0 & \ldots & 0 & 1 \\
0 & 0 & 0 & \ldots & 0 & c & b
\end{bmatrix} \in M_{2k}(\mathbb{R})$. 

Exercises

In the exercises, $K$ is a field, $V$ is a finite dimensional $K$-vector space and $u$ is an endomorphism of $K V$.

1. Prove that the eigenvectors corresponding to distinct eigenvalues of $u$ are linearly independent.

2. Give an example of three matrices in $M_3(\mathbb{Q})$ whose only eigenvalue is 2 and any two matrices are not similar. Can you exhibit four such matrices? Generalization.

3. Determine the endomorphisms $u \in \text{End}_K(V)$ whose minimal polynomials are of degree 1.

4. Give an example of two matrices having the same minimal polynomial and the same characteristic polynomial, but are not similar.

5. Let $A \in M_n(K)$ such that the characteristic polynomial of $A$ splits in factors of degree 1 in $K[X]$ (one says that $A$ has all its eigenvalues in $K$). Then $A$ is similar to an upper triangular matrix $T = (t_{ij}) \in M_n(K)$ ($t_{ij} = 0$ if $i > j$). In this case, $A$ is called trigonable. Is the converse true?

6. Let $A \in M_n(K)$ and let $p \in K[X]$. If $A$ has all its eigenvalues $\lambda_1, \ldots, \lambda_n$ in $K$, then $p(A)$ has the eigenvalues $p(\lambda_1), \ldots, p(\lambda_n)$. (Ind. Let $A \approx T$, with $T$ upper triangular. The diagonal of $T$ is $\lambda_1, \ldots, \lambda_n$. Compute $p(T)$.)

7. Compute the characteristic polynomial and the minimal polynomial of the following matrices:

\[
\begin{bmatrix}
1 & 0 & 1 & -2 \\
0 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 \\
0 & 0 & 0 & 2
\end{bmatrix},
\begin{bmatrix}
-7 & 3 & 3 \\
-21 & 9 & 7 \\
-6 & 2 & 4
\end{bmatrix},
\begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

8. Let $A \in M_n(K)$. Then $^tA$ is similar to $A$. Moreover, there exists $U \in \text{GL}(n, K)$, symmetric, such that $^tA = U^{-1}AU$. 

III.4 The endomorphisms of a finite dimensional vector space
9. Let $R$ be a commutative ring with identity and let $n \in \mathbb{N}^*$. Generalize the relevant notions and prove the Cayley-Hamilton theorem: if $E$ is a free $R$-module of rank $n$ and $u \in \text{End}_R(E)$, then $u$ is a root of the characteristic polynomial: $f_u(u) = 0$. (Hint: let $A$ the matrix of $u$ in a basis $(e_1, \ldots, e_n)$ and $XI - A = B = (b_{ij}) \in M_n(R[X])$; $f_u = \text{det} B$. In the $R[X]$-module $E_u$ the relations $\sum_j b_{ij} e_j = 0$ hold, $\forall i$. Let $B_{ik} \in R[X]$ be the algebraic complement of $b_{ik}$ in the matrix $B$. For any fixed $k$, multiply relation $i$ with $B_{ik}$ and sum after $i$ to obtain $f_u(X) \cdot e_k = 0 = f_u(u)(e_k)$.)

10. Let $R$ be a commutative ring with identity and let $A \in U(M_n(R))$ be an invertible matrix. Then $A$ is a $R$-linear combination of $I, A, \ldots, A^{n-1}$. (Hint. Use the Cayley-Hamilton theorem.)

11. Let $u \in \text{End}_K(V)$. Then $V$ has no $u$-invariant proper subspaces $\iff$ the $K[X]$-module $V_u$ is simple $\iff$ the characteristic polynomial of $u$ is irreducible in $K[X]$.

12. Let $u \in \text{End}_K(V)$ be an endomorphism having the eigenvalue $0$. Then $V = U \oplus W$, with $U, W u$-invariant subspaces and dim $U = 1$.

13. Let $u \in \text{End}_K(V)$ be a nilpotent endomorphism ($\exists r \geq 1$ such that $u^r = 0$). Then $\text{Tr} u^k = 0$, $\forall k \geq 1$. Conversely, if char $K = 0$ and $\text{Tr} u^k = 0$, $\forall k \geq 1$, then $u$ is nilpotent. (Ind. Let $f$ be the characteristic polynomial of $u$; in the relation $f(u) = 0$ apply Tr and deduce that 0 is an eigenvalue of $u$. So, $u$ has an invariant subspace of dimension $\dim V - 1$.)

14. Let $V = U \oplus W$ (direct sum of subspaces). Then any $v \in V$ can be uniquely written as $v = u + w$, with $u \in U, \ w \in W$. Define $\pi, \rho: V \to V$ by: $\forall u \in U, \forall w \in W, \pi(u + w) = u$ ($\pi$ is called the projection on $U$ along $W$) and $\rho(u + w) = u - w$, ($\rho$ is called the symmetry with respect to $U$ along $W$). Show that $\pi$ and $\rho$ are $K$-endomorphisms of $V$ and find their minimal polynomials.
IV. Field extensions

This chapter contains the basic concepts and results from the theory of field extensions. Standard facts about rings and vector spaces are a prerequisite: polynomial rings, factor rings, ring isomorphism theorems, bases and dimension in vector spaces, prime and maximal ideals. Knowledge of polynomial ring arithmetic is recommended (as provided in the chapter “Arithmetic in integral domains”). Some elementary properties of cardinals and Zorn's Lemma are used in the proof of existence of the algebraic closure of a field. Most of these facts can be found in the Appendices; a more detailed treatment is found in most Abstract (Modern) Algebra introductory texts.

IV.1. Algebraic extensions

Recall that a field is a commutative ring \((K, +, \cdot)\) with identity 1 (with \(1 \neq 0\)), with the property that every nonzero element is invertible with respect to multiplication. Any field has at least two elements: 0 and 1. A field \(K\) has no zero divisors (in other words, \(K\) is a domain): for any \(x, y \in K, x \neq 0\) and \(y \neq 0\) implies \(xy \neq 0\).
All rings and all ring homomorphisms considered are supposed to be unitary. Thus, if $K, L$ are rings (with identity), the map $\sigma: K \to L$ is a homomorphism if and only if, $\forall x, y \in K$:

$$\sigma(x + y) = \sigma(x) + \sigma(y)$$
$$\sigma(xy) = \sigma(x)\sigma(y)$$
$$\sigma(1) = 1.$$ 

The objects we study are field extensions: if $L$ is a field and $K$ is a subfield in $L$, we also say that “$L$ is an extension of $K$”.

Recall that the nonempty set $K$ of the field $L$ is a subfield in $L$ if it is closed under addition and multiplication and becomes a field with the induced operations. A widely used characterization is:

$K$ is a subfield of $L$ if and only if for any $x, y \in K$ with $y \neq 0$, we have $x \cdot y$ and $xy^{-1} \in K$.

The subfield $K$ of $L$ is called proper if $K \neq L$.

This notion of field extension is too restrictive. For instance, it is natural to consider that $\mathbb{C}$ is an extension of $\mathbb{R}$. But $\mathbb{C}$ is usually constructed as the set of all couples $(a, b)$ with $a, b \in \mathbb{R}$, endowed with an addition and a multiplication that make it a field. In this setting, $\mathbb{R}$ is not a subset of $\mathbb{C}$, but $\mathbb{R}$ can be identified with the set of the couples $(a, 0), a \in \mathbb{R}$. In fact, the rigorous interpretation is the following: one defines the field homomorphism $\varphi: \mathbb{R} \to \mathbb{C}, \varphi(a) = (a, 0)$ and then identifies $\mathbb{R}$ with its image $\varphi(\mathbb{R})$, which is a subfield of $\mathbb{C}$.

More generally, if $\sigma: K \to L$ is a field homomorphism, then $\sigma$ is injective. Indeed, $\forall x \in K, x \neq 0$ implies $\exists x^{-1} \in K$, so $\sigma(x)\sigma(x^{-1}) = \sigma(xx^{-1}) = \sigma(1) = 1$, so $\sigma(x)$ is nonzero and thus $\ker \sigma$ is $\{0\}$. The homomorphism $\sigma: K \to L$ being injective, we may identify the field $K$ with its image $\sigma(K)$, which is a subfield in $L$. It is thus natural to consider the following definition:

1.1 Definition. a) Let $K, L$ be fields. If $\sigma: K \to L$ is a field homomorphism, we call the triple $(K, L, \sigma)$ a field extension of $K$. We ex-
press this by writing “$K \subseteq L$ is a field extension”, “$L/K$ is a field extension” or “$L$ is an extension of $K$”. For any element $a \in K$, we identify $\sigma(a) \in L$ with $a \in K$. For example, if $a \in K$ and $x \in L$, we write $a \cdot x$ instead of $\sigma(a) \cdot x$. This identification allows to consider $K$ as a subfield of $L$. The extension $K \subseteq L$ is called proper if the inclusion is strict (i.e. $\sigma(K) \subset L$). In this context, we call an intermediate field (or subextension) of the extension $K \subseteq L$ any subfield $E$ of $L$ that includes $K$. The intermediate field $E$ is called proper if $E \neq K$ and $E \neq L$.

Note that $L$ is an extension of $K$ if and only if $L$ is a field also structured as a $K$-algebra (the structural homomorphism is precisely $\sigma$).

b) If $K \subseteq L$ and $K \subseteq E$ are extensions of $K$, a mapping $\varphi : L \to E$ is called a $K$-homomorphism if $\varphi$ is a ring homomorphism and $\varphi|_K = \text{id}_K$ (the identity function of $K$).

If one considers the general definition above, i.e. there exist field homomorphisms $\sigma : K \to L$, $\tau : K \to E$, we call $\varphi : L \to E$ a $K$-homomorphism if $\varphi$ is a field homomorphism with $\varphi \circ \sigma = \tau$ (in other words, $\varphi$ is a $K$-algebra homomorphism). Let $\text{Hom}_K(L, E)$ denote the set of all $K$-homomorphism from $L$ to $E$.

1.2 Examples. a) The real numbers field $\mathbb{R}$ is an extension of $\mathbb{Q}$, the field of rational numbers.

b) The field $\mathbb{C}$ of complex numbers is an extension of $\mathbb{R}$.

c) The set $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \in \mathbb{C} \mid a, b \in \mathbb{Q}\}$ is a field, and it is an extension of $\mathbb{Q}$.

d) If $K$ is a field, then $K(\mathbb{X})$, the field of rational fractions with coefficients in $K$ ($K(\mathbb{X})$ is the quotient field of the polynomial ring $K[\mathbb{X}]$) is an extension of $K$.

e) If $n$ is a natural number, the ring $\mathbb{Z}_n$ of the integers modulo $n$ is a field if and only if $n$ is a prime. For any prime $p$, let $\mathbb{F}_p$ denote the field of integers modulo $p$. We shall construct an extension of $\mathbb{F}_2$ at 1.24.
A method to construct subfields, very important in Galois theory\textsuperscript{39}, is the following: let $K$ be a field and let $H$ be a set of field endomorphisms of $K$ (homomorphisms defined on $K$ with values in $K$). The set $K^H := \{ x \in K | \sigma(x) = x, \forall \sigma \in H \}$ is a subfield of $K$ (called the \textit{fixed subfield of} $H$). Indeed, if $x, y \in K^H$ and $\sigma \in \text{End}(K)$ then $\sigma(x - y) = \sigma(x) - \sigma(y) = x - y$, so $x - y \in K^H$. Likewise, $xy^{-1} \in K^H$ if $y \neq 0$.

A basic tool is the concept of \textit{degree} of an extension.

\textbf{1.3 Definition.} If $K \subseteq L$ is an extension, then $L$ is canonically a $K$-vector space: the multiplication of a “scalar” in $K$ with a “vector” in $L$ is their multiplication in $L$. The dimension of $L$ as a $K$-vector space is called the \textit{degree}\textsuperscript{40} of the extension $K \subseteq L$ and is denoted by $[L : K]$ or $(L : K)$. An extension is called \textit{finite} if its degree is finite. In example b) above, $\{1, i\}$ is a basis of $\mathbb{C}$ over $\mathbb{R}$, so $[\mathbb{C} : \mathbb{R}] = 2$. What are the degrees of the other extensions?

\textbf{1.4 Definition.} The fields that have no proper subfields are called \textit{prime fields}.

We determine now all \textit{prime fields}.

Recall the notion of \textit{characteristic} of a ring $R$ with identity. Let $1$ be the identity element of $R$ and let $n \in \mathbb{N}$; $n \cdot 1$ denotes the \textit{multiple} $1 + \ldots + 1$ (n terms); the \textit{characteristic} of $R$, denoted $\text{char} \ R$, is defined as follows:

- if, for any $n \in \mathbb{N}^*$, $n \cdot 1 \neq 0$, then $\text{char} \ R = 0$;

---

\textsuperscript{39} Evariste Galois (1811-1832), French mathematician.

\textsuperscript{40} One can define the degree of an extension of skew fields (division rings): if $L$ is a skew field and $K$ is a subfield (skew) of $L$, then $L$ is naturally a $K$-vector space and one defines the degree $[L : K]$ as $\dim_K L$. 
- if there exists \( n \in \mathbb{N}^* \) such that \( n \cdot 1 = 0 \), then \( \text{char} \ R \) is the smallest \( n \in \mathbb{N}^* \) such that \( n \cdot 1 = 0 \).

1.5 Remark. The characteristic of \( R \) can be defined also as follows: there exists a unique ring homomorphism \( \varphi : \mathbb{Z} \to R \) (prove!). Then \( \text{char} \ R \) is the natural generator of the ideal \( \text{Ker} \varphi \). The proof of the equivalence of these definitions is left to the reader.

For example, \( \text{char} \ \mathbb{Z} = \text{char} \ \mathbb{Q} = 0 \); \( \text{char} \ \mathbb{F}_2 = 2 \); if \( R \) is a unitary subring of \( S \), then \( \text{char} \ R = \text{char} \ S \).

1.6 Proposition. Let \( R \) be a domain. Then the characteristic of \( R \) is 0 or a prime. In particular, the characteristic of a field is 0 or a prime.

Proof. Suppose \( n = \text{char} \ R \neq 0 \). If, by contradiction, \( n = ab \), with \( a, b \in \mathbb{N}^* \), \( a < n, b < n \), then \( 0 = n \cdot 1 = (a \cdot 1)(b \cdot 1) \). But \( R \) has no zero divisors, so \( a \cdot 1 = 0 \) or \( b \cdot 1 = 0 \), contradicting the minimality of \( n \).

A related notion is the characteristic exponent of a field: the characteristic exponent of the field \( K \) is 1 if \( \text{char} \ K = 0 \) and is \( p \) if \( \text{char} \ K = p > 0 \).

1.7 Proposition. Let \( K \) be a prime field. If \( \text{char} \ K = 0 \), then there exists a unique isomorphism \( K \cong \mathbb{Q} \). If \( \text{char} \ K = p > 0 \), then there exists a unique isomorphism \( K \cong \mathbb{F}_p \).

Proof. Define the ring homomorphism \( \varphi : \mathbb{Z} \to K, \varphi(n) = n \cdot 1, \forall n \in \mathbb{Z} \), where 1 is the identity of \( K \).

Suppose \( \text{char} \ K = 0 \). Since for any \( n \in \mathbb{Z}, n \cdot 1 \neq 0 \), \( n \cdot 1 \) is invertible in \( K \). Define the homomorphism \( \psi : \mathbb{Q} \to K \) that extends \( \varphi \), \( \psi(a/b) = \varphi(a)\varphi(b)^{-1}, \forall a, b \in \mathbb{Z}, b \neq 0 \). Since \( \psi(\mathbb{Q}) \) is a subfield in \( K \), we have \( \psi(\mathbb{Q}) = K \), so \( \psi \) is an isomorphism.

If \( \text{char} \ K = p \), \( \text{Ker} \varphi = p\mathbb{Z} \); applying the isomorphism theorem, \( \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p \cong \text{Im} \varphi \), which is a subfield in \( K \). Since \( K \) has no proper subfields, \( \text{Im} \varphi = K \).

The reader is invited to prove the uniqueness part.
1.8 Lemma. The intersection of any family of subfields of a field is a subfield.

Proof. Use the characterization: $K$ is a subfield of $L$ if and only if for any $x, y \in K$ with $y \neq 0$, we have $x - y$ and $xy^{-1} \in K$. If $(L_i)_{i \in I}$ is a family of subfields of the field $L$ and $x, y \in \bigcap L_i$, then $x, y \in L_i, \forall i \in I$, so $x - y \in L_i, \forall i \in I$, i.e. $x - y \in \bigcap L_i$. The same argument works for $xy^{-1}$ if $y \neq 0$.

1.9 Theorem. Any field $K$ contains a unique prime subfield. In particular:
- if $\text{char } K = 0$, then $K$ is an extension of $\mathbb{Q}$;
- if $\text{char } K = p > 0$, then $K$ is an extension of $\mathbb{F}_p$.

Proof. The intersection of all subfields of $K$ is a subfield $P$, which cannot have proper subfields (any subfield $F$ of $P$ would be a subfield in $K$, so $F$ includes $P$, absurd!). If $Q$ is a prime subfield of $K$, then $P \cap Q$ is a subfield both in $P$ and in $Q$, so it is equal to both of them; this shows that $P$ is the unique prime subfield in $K$. The rest of the statement follows from the list of prime subfields.

1.10 Application. Finite fields. If $F$ is a field with a finite number of elements, then $\text{char } F$ cannot be 0 (otherwise $F$ is an extension of $\mathbb{Q}$, which is infinite!), so $\text{char } F = p > 0$ (for some prime $p$) and $F$ is an extension of the field $\mathbb{F}_p$. If $n$ is the degree $[F : \mathbb{F}_p]$, $n$ is finite and $F$ is isomorphic (as an $\mathbb{F}_p$-vector space) to $(\mathbb{F}_p)^n$. Thus, the cardinal of a finite field is $p^n$, for some prime $p$ and some $n \in \mathbb{N}^*$. This means there do not exist fields with 10 elements, for instance.

The following definition describes a basic construction in field extensions.

1.11 Definition. Let $K \subseteq L$ be a field extension and let $S$ be a subset of $L$. The intersection of all subfields of $L$ that include $K \cup S$ is a subfield of $L$, called the field generated by $K$ and $S$, and denoted by
$K(S)$. One also says that $K(S)$ is obtained by \textit{adjoining} to $K$ the elements in $S$.

We denote by $K[S]$ the \textit{subring of $L$ generated by $K \cup S$} (the intersection of all subrings of $L$ that include $K \cup S$).\footnote{This notation is used also for ring extensions: if $R$ is a subring of the ring $A$, and $S$ is a subset of $A$, $R[S]$ is the subring generated by $R \cup S$ in $A$ ($R[S]$ coincides with the $R$-subalgebra of $A$ generated by $S$).}

It is easy to see that $K(S)$ (respectively $K[S]$) is the smallest subfield (respectively subring) of $L$ that includes $K \cup S$. Obviously, the ring $K[S]$ is a domain (it is a subring of the field $L$) and $K[S] \subseteq K(S)$. The field of fractions of the domain $K[S]$ is canonically isomorphic to $K(S)$ (see the universality property of the field of fractions).

If $S = \{x_1, \ldots, x_n\}$, then $K(S)$ is denoted by $K(x_1, \ldots, x_n)$ and $K[S]$ by $K[x_1, \ldots, x_n]$. An extension $K \subseteq L$ with the property there exists a finite subset $S$ of $L$ such that $L = K(S)$ is called a \textit{finitely generated extension}. Do not confuse with the concept of \textit{finite extension} (which means that its degree is finite)!

If there exists $\alpha \in L$ such that $L = K(\alpha)$, the extension $K \subseteq L$ is called a \textit{simple extension}, and $\alpha$ is called a \textit{primitive element}. A primitive element need not be unique: for instance, $K(\alpha) = K(\alpha + 1)$.

\textbf{1.12 Definition.} Let $K \subseteq L$ be a field extension and let $E$, $F$ be extensions of $K$, included in $L$. The \textit{composite} of the fields $E$ and $F$ (denoted by $EF$) is the subfield of $L$ generated by $E$ and $F$: $EF = K(E \cup F) = E(F) = F(E)$. In general, for a family $(E_i)_{i \in I}$ of extensions of $K$, included in $L$, the \textit{composite of the fields $(E_i)_{i \in I}$} is the subfield of $L$ generated by $\bigcup_{i \in I} E_i$, $K(\bigcup_{i \in I} E_i)$.  

1.13 Remark. The set \( \text{IF}(L/K) \) of all intermediate fields of a given extension \( L/K \) is ordered by inclusion and \( \text{IF}(L/K) \) is a (complete) lattice\(^{42} \): for any \( E, F \in \text{IF}(L/K) \), \( \sup\{E, F\} = EF \), \( \inf\{E, F\} = E \cap F \).

1.14 Theorem. Let \( L/K \) be an extension and let \( S, S_1, S_2 \) be subsets of \( L \). Then:

a) \( K[S] \) is the set of all polynomial expression in the elements of \( S \), with coefficients in \( K \), namely:

\[
K[S] = \left\{ \sum_{(i_1, \ldots, i_n) \in \mathbb{N}^n} a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \middle| n \in \mathbb{N}^+, x_1, \ldots, x_n \in S, a_{i_1 \ldots i_n} \in K, \forall (i_1, \ldots, i_n) \in \mathbb{N}^n \right\}
\]

where \( \sum' \) means the sums are finite, i.e. the set \( \{(i_1, \ldots, i_n) \in \mathbb{N}^n | a_{i_1 \ldots i_n} \neq 0 \} \) is finite.

b) \( K(S) = \{ \alpha \beta^{-1} \middle| \alpha, \beta \in K[S], \beta \neq 0 \} \).

c) \( K[S_1 \cup S_2] = K[S_1][S_2] = K[S_2][S_1] \) and \( K(S_1 \cup S_2) = K(S_1)(S_2) = K(S_2)(S_1) \).

d) \( K(S) = \bigcup \{ K(T) \mid T \subseteq S, T \text{ finite} \} \).

Proof. a) Let \( T \) be the set of all polynomial expression in the elements of \( S \), with coefficients in \( K \). One checks directly that \( T \) is a subring of \( L \), that includes \( K \) and \( S \). (Alternatively, \( T \) is the image in \( L \) via the evaluation homomorphism of polynomial ring in \( S \) indeterminates). On the other hand, any subring of \( L \) that includes \( K \) and \( S \) must also include \( T \). Thus, \( T \) is the smallest subring of \( L \) that includes \( K \) and \( S \).

---

\(^{42}\) An ordered set \((S, \leq)\) is called a lattice (resp. a complete lattice) if any subset with two elements (resp. any subset) of \( S \) has a least upper bound and a greatest lower bound in \( L \).
b) Since $K[S] \subseteq K(S)$ and $K(S)$ is a field, any element of the form $\alpha\beta^{-1}$, with $\alpha, \beta \in K[S]$, $\beta \neq 0$, is in $K(S)$. But the set of all these elements is a subfield in $L$ (standard check).

c), d) Exercise.

In the particular case $S = \{x_1, \ldots, x_n\}$, with $x_1, \ldots, x_n \in L$, :

$$K[x_1, \ldots, x_n] = \{f(x_1, \ldots, x_n) \mid f \in K[X_1, \ldots, X_n]\}$$

$$K(x_1, \ldots, x_n) = \left\{ \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} \mid f, g \in K[X_1, \ldots, X_n], g(x_1, \ldots, x_n) \neq 0 \right\}.$$

If $S = \{a\} \in L$:

$K[a] = \{f(a) \mid f \in K[X]\}$ and $K(a) = \{f(a)/g(a) \mid f, g \in K[X], g(a) \neq 0\}$.

Thus, $K[a] = \text{Im } ev_a$, where $ev_a : K[X] \to L$ is the unique $K$-algebra homomorphism with the property that $ev_a(a) = a$; $ev_a$ is called the “homomorphism of evaluation in $a$”. If $f \in K[X], f = b_0 + b_1 X + \ldots + b_n X^n$, then $ev_a(f) = b_0 + b_1 a + \ldots + b_n a^n \in L$. The usual notation for $ev_a(f)$ is $f(a)$, called the “value of $f$ in $a$”. If $f(a) = 0$, we say „$a$ is a root of $f$”.

The following notion is central in all the theory we describe.

1.15 Definition. Let $K \subseteq L$ be a field extension and let $x \in L$. We say that is algebraic over $K$ if there exists a nonzero polynomial $f \in K[X]$ such that $f(x) = 0$.

In other words, $x$ is algebraic over $K$ if and only if the evaluation homomorphism $ev_x : K[X] \to L$ is not injective.

If the element $x$ is not algebraic over $K$ (i.e. $ev_x$ is injective), $x$ is called transcendental over $K$. Thus, $x$ is transcendental over $K$ if and only if $ev_x$ induces a $K$-isomorphism between the polynomial ring $K[X]$ and the subring $K[x]$ generated by $K$ and $x$ in $L$.

1.16 Examples. a) In the extension $\mathbb{Q} \subseteq \mathbb{R}$, the element $\sqrt{2}$ is algebraic over $\mathbb{Q}$, as a root of $X^2 - 2 \in \mathbb{Q}[X]$. 
b) Any element of $K$ is algebraic over $K$.

c) Let $K(X)$ be the field of rational fractions in the indeterminate $X$ with coefficients in the field $K$. The element $X$ in the extension $K \subseteq K(X)$ is transcendental over $K$. Indeed, the evaluation homomorphism $\text{ev}_X : K[X] \rightarrow K(X)$ is the canonical inclusion, and is injective.

However, $X$ is obviously algebraic over $K(X)$. This underscores the importance of identifying the field over which the element is algebraic.

d) A complex number that is algebraic (respectively transcendental) over $\mathbb{Q}$ is called by tradition algebraic number (respectively transcendental number), without any reference to $\mathbb{Q}$.

e) The real numbers $e$ (the sum of the series $\sum_{n \geq 0} (n!)^{-1}$) and $\pi$ (the length of a circle of diameter 1) are transcendental (over $\mathbb{Q}$). These facts were proven by Hermite in 1873 and Lindemann in 1882.

With the notations above, $x$ is algebraic over $K$ if and only if $\text{Ker} \text{ev}_x \neq \{0\}$. The generator of the ideal $\text{Ker} \text{ev}_x$ ($K[X]$ is a PID!) is a nonzero polynomial, which is fundamental in describing the extension $K \subseteq K(x)$.

1.17 Theorem. Let $K \subseteq L$ be a field extension and let $x \in L$, algebraic over $K$. Let $f$ be a monic polynomial with coefficients in $K$. The following statements are equivalent:

a) $f(x) = 0$ and $\deg f$ is minimal among the nonzero polynomials that vanish in $x$:
\[
\deg f = \min \{\deg g \mid g \in K[X], g(x) = 0, g \neq 0\}.
\]

b) $f(x) = 0$ and $f$ is irreducible.

c) $f$ is a generator of the ideal $\text{Ker} \text{ev}_x = \{g \in K[X] \mid g(x) = 0\}$.

d) $f(x) = 0$ and, for any $g \in K[X]$, $g(x) = 0$ implies $f \mid g$.

Proof. a)$\implies$b) If $f$ were reducible, then $f = gh$, for some $g, h \in K[X]$, with $1 \leq \deg h$, $\deg g < \deg f$. Since $g(x)h(x) = f(x) = 0$, $x$ is a root of $g$
or of \( h \), whose degrees are less than \( \deg f \), contradicting the hypothesis.

\( b) \Rightarrow c) \) Clearly, \( (f) \subseteq \text{Ker } ev_x \). In the PID \( K[X] \), the ideal generated by \( f \) is maximal, because \( f \) is irreducible. So, \( \text{Ker } ev_x = (f) \) or \( \text{Ker } ev_x = K[X] \). But obviously \( 1 \notin \text{Ker } ev_x \iff \text{Ker } ev_x \neq K[X] \), so \( (f) = \text{Ker } ev_x \).

\( c) \Leftrightarrow d) \) is evident.

\( d) \Rightarrow a) \) Let \( g \in K[X] \) with \( g(x) = 0 \), \( g \neq 0 \). By the hypothesis, \( f \mid g \), so \( \deg f \leq \deg g \).

1.18 Definition. Let \( L/K \) be a field extension and let \( x \in L \) be algebraic over \( K \). We call minimal polynomial of \( x \) over \( K \) the monic polynomial in \( K[X] \) that satisfies one of the equivalent properties in the previous proposition. The minimal polynomial of \( x \) over \( K \) is denoted by \( \text{Irr}(x, K) \) or \( \text{min}(x, K) \).

1.19 Examples. a) \( \text{Irr}(\sqrt[4]{2}, \mathbb{Q}) = X^4 - 2 \) since \( X^4 - 2 \in \mathbb{Q}[X] \) is monic, \( \sqrt[4]{2} \) is a root and it is irreducible in \( \mathbb{Q}[X] \) (by Eisenstein’s criterion).

b) \( \text{Irr}(\sqrt[4]{2}, \mathbb{R}) = X - \sqrt[4]{2} \). More generally, for any field \( K \) and any \( a \in K \), \( \text{Irr}(a, K) = X - a \).

c) \( \text{Irr}(\sqrt[4]{2}, \mathbb{Q}(\sqrt[2]{2})) = X^2 - \sqrt{2} \). The fact \( X^2 - \sqrt{2} \) that has minimum degree among the polynomials with coefficients in \( \mathbb{Q}(\sqrt[2]{2}) \) that vanish in \( \sqrt[4]{2} \) is equivalent to \( \sqrt[4]{2} \notin \mathbb{Q}(\sqrt[2]{2}) \). (prove this, using the form of the elements in \( \mathbb{Q}(\sqrt[2]{2}) \)).

1.20 Theorem (characterization of algebraic elements in an extension). Let \( K \subseteq L \) be a field extension and let \( x \in L \). The following statements are equivalent:

\( a) x \) is algebraic over \( K \).

\( b) K[x] \) is a field.

\( c) K[x] = K(x) \).
d) The extension $K \subseteq K(x)$ is finite.

Besides, if $x$ is algebraic over $K$, then the degree of the extension $K \subseteq K(x)$ is the degree of the minimal polynomial $\text{Irr}(x, K)$:

$$[K(x) : K] = \deg \text{Irr}(x, K)$$

If $n = \deg f$, a $K$-basis of the extension $K(x)$ is $\{1, x, \ldots, x^{n-1}\}$.

**Proof.** $a) \Rightarrow b)$ Let $f = \text{Irr}(x, K) \in K[X]$ and let $\text{ev}_x : K[X] \to L$ be the evaluation homomorphism in $x$. The ideal $\text{Ker} \text{ev}_x$ of $K[X]$ is generated by $f$. The isomorphism theorem implies $K[X]/(f) \cong \text{Im} \text{ev}_x = K[x]$. Since $f$ is irreducible in $K[X]$, the ideal $(f)$ is maximal and $K[X]/(f)$ is a field. Then $K[x]$, isomorphic to $K[X]/(f)$, is also a field.

It is evident the $b) \iff c)$.

$c) \Rightarrow a)$ If $x = 0$, then all is evident. Suppose $x \neq 0$ and let $x^{-1} = a_0 + a_1x + \ldots + a_nx^n \in K[x]$ be the inverse of $x$. Multiplying by $x$, we get $a_0 + a_1x^2 + \ldots + a_nx^{n+1} - 1 = 0$, so $x$ is a root of a nonzero polynomial with coefficients in $K$.

$d) \Rightarrow a)$ The infinite family $\{x^i \mid i \in \mathbb{N}\}$ of elements of the finite dimensional $K$-vector space $K(x)$ is linearly dependent. So, there exists a linear dependence relation, $a_01 + a_1x + \ldots + a_nx^n = 0$, for some $n \in \mathbb{N}$ and $a_0, a_1, \ldots, a_n \in K$, not all zero. This shows that $x$ is algebraic over $K$.

$a) \Rightarrow d)$ At $a) \Rightarrow b)$ we saw that we have a $K$-isomorphism of fields $K[X]/(f) \cong K(x)$. Evidently, this is also a isomorphism of $K$-vector spaces. Let $n = \deg f$. In the $K$-vector space $K[X]/(f)$, the classes of $1, X, \ldots, X^{n-1}$ form a basis. If

$$a_0(1 + (f)) + a_1(X + (f)) + \ldots + a_{n-1}(X^{n-1} + (f)) = 0 + (f),$$

where $a_0, a_1, \ldots, a_{n-1} \in K$, then $g = a_0 + a_1X + \ldots + a_{n-1}X^{n-1} \in (f)$, so $f | g$. Since $\deg f = n$, we have $g = 0$, i.e. $a_0, a_1, \ldots, a_{n-1}$ are 0. On the other hand, the division with remainder theorem shows that any class modulo $f$ of a polynomial $h \in K[X]$ has a un representative of degree less than $n$. This means that $h + (f)$ is a linear combination with coefficients in $K$ of the classes of $1, X, \ldots, X^{n-1}$.
IV.1. Algebraic extensions

So, \( \dim_K K(x) = \dim_K K[X]/(f) = n. \)

The isomorphism \( K[X]/(f) \cong K(x) \) takes the basis \( 1 + (f), X + (f), \ldots, X^{n-1} + (f) \) in the basis \( 1, x, \ldots, x^{n-1}. \)

1.21 Remark. Keep the conventions above. The proof of \( a) \Rightarrow b) \) above uses essentially the fact that if \( f \) is irreducible in \( K[X] \), then \( K[X]/(f) \) is a field. This result can be proven in a more “elementary” way, as follows. We must show that any nonzero element in \( K[X]/(f) \) has a multiplicative inverse. Take \( g \in K[X], \ g + (f) \neq 0 + (f). \) This implies \( f \mid g, \) so \( \gcd(f, g) = 1. \) By the extended Euclid algorithm, there exist (and can be effectively computed) \( u, v \in K[X] \) such that \( uf + vg = 1. \) Taking classes modulo \( (f), \) this relation shows that \( v + (f) \) is the inverse of \( g + (f) \) in \( K[X]/(f). \)

This also suggests a way to compute the inverse of an arbitrary element in \( K[x] \) expressed in the basis \( \{1, x, \ldots, x^{n-1}\}: \) for any element \( y = a_0 + a_1 X + \ldots + a_{n-1} X^{n-1} =: g(x), \) where \( g \in K[X], \ g \neq 0, \) \( \deg g < n, \) then \( y^{-1} \) is \( v(x), \) where \( v, u \in K[X] \) are such that \( uf + vg = 1. \)

Let us find, in the extension \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \), the inverse of \( y = 1 + \sqrt[3]{2} + \sqrt[3]{4}. \) Let \( x = \sqrt[3]{2}. \) We have \( \operatorname{Irr}(x, K) = f = X^3 - 2 \) and \( y = g(\sqrt[3]{2}), \) where \( g = 1 + X + X^2. \) Using the extended Euclid algorithm for \( X^3 - 2 \) and \( 1 + X + X^2, \) we find that \( 1 = ( - 1)(X^3 - 2) + (X - 1)(1 + X + X^2). \) Evaluating in \( \sqrt[3]{2}, \) we obtain \( 1 = (\sqrt[3]{2} - 1)(1 + \sqrt[3]{2} + \sqrt[3]{4}). \)

1.22 Definition. Let \( K \) be a field and let \( x \) be algebraic over \( K. \) The degree of the extension \( K \subseteq K[x] \) (equal to \( \deg \operatorname{Irr}(x, K) \)) is called the degree of \( x \) over \( K. \)

The factor ring \( K[X]/(f) \) appears also in the following proof.
1.23 Proposition. Let \( K \) be a field and let \( f \in K[X] \), \( \deg f \geq 1 \). Then there exists an extension of \( K \) in which \( f \) has a root.

Proof. In the UFD \( K[X] \), the polynomial \( f \) is a product of irreducible polynomials. Any root of a factor of \( f \) is also a root of \( f \). Replacing, if needed, \( f \) with one of its irreducible factors, it is sufficient to prove the claim for the case when \( f \) is irreducible.

Consider the factor ring \( K[X]/\langle f \rangle = L \), which is a field, since \( f \) is irreducible. Moreover, \( L \) is an extension of \( K \), because the canonical mapping \( \varphi : K \to K[X]/\langle f \rangle \), \( \varphi(a) = a + \langle f \rangle \), \( \forall a \in K \), is a field homomorphism. In \( L \), \( \alpha = X + \langle f \rangle \) is a root of \( f \). Indeed, suppose that

\[
    f(\alpha) = a_0 + a_1X + \ldots + a_nX^n
\]

Then \( f(\alpha) = a_0(1 + \langle f \rangle) + a_1(X + \langle f \rangle) + \ldots + a_n(X^n + \langle f \rangle) = (a_0 + a_1X + \ldots + a_nX^n) + \langle f \rangle = f + \langle f \rangle = 0 + \langle f \rangle. \]

1.24 Examples. The importance of this result is not just theoretical. Concrete fields can be constructed following the procedure in the proof. Here are two examples.

a) Let \( f = X^2 + 1 \in \mathbb{R}[X] \). The polynomial \( f \) is irreducible in \( \mathbb{R}[X] \): it has degree 2 and has no roots in \( \mathbb{R} \). So, \( \mathbb{R}[X]/\langle X^2 + 1 \rangle \) is a field, extension of \( \mathbb{R} \). Let \( \bar{g} \) be the class of the polynomial \( g \in \mathbb{R}[X] \) modulo \( \langle X^2 + 1 \rangle \). The class of \( X \) modulo \( \langle X^2 + 1 \rangle \), denoted by \( i \), is a root of \( f \). Any element of this field is written uniquely as the class of a polynomial of degree at most 1 modulo \( \langle X^2 + 1 \rangle \), i.e. is of the form

\[
    a+bX = \bar{a} + \bar{b} \cdot \bar{X} = \bar{a} + \bar{b} \cdot i,
\]

where \( a, b \in \mathbb{R} \). Identifying the real

\[43\] This result, whose proof is simple for a modern mathematician, was known intuitively for a long time and often tacitly accepted in 17th to 19th centuries’ mathematical arguments (of course, only polynomials with numerical coefficients were considered). Around 1629, Albert Girard, states –without proof– that an equation of degree \( n \) has \( n \) roots, that can be complex numbers or “other similar numbers”. In 1792, Pierre Simon de Laplace gives an elegant proof of the “Fundamental Theorem of Algebra” – any nonconstant polynomial with complex coefficients has a complex root – admitting though that the roots exist “somewhere”.
number \( a \) with its class \( \bar{a} \), we can uniquely write a generic element of \( \mathbb{R}[X]/(X^2 + 1) \) in the form \( a + bi \), with \( a, b \in \mathbb{R} \).

In this field we have \( i^2 = -1 \). Thus, the multiplication of two elements is given by:

\[
(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i, \quad \forall a, b, c, d \in \mathbb{R}
\]

This field is isomorphic to the complex number field \( \mathbb{C} \). In fact, this construction can be taken as a definition of \( \mathbb{C} \).

b) **Construction of a field with \( p^n \) elements.** Let \( p \) be a prime number and let \( n \in \mathbb{N}^* \). If \( f \in \mathbb{F}_p[X] \) is an irreducible polynomial of degree \( n \), then \( \mathbb{F}_p[X]/\langle f \rangle \) is a field with \( p^n \) elements.

For instance, the polynomial \( f = X^2 + X + 1 \) is irreducible in \( \mathbb{F}_2[X] \) (it has degree 2 and has no roots in \( \mathbb{F}_2 \)). So, \( K := \mathbb{F}_2[X]/\langle f \rangle \) is a field with 4 elements. If \( \alpha \) denotes the class of \( X \) modulo \( f \), any element in \( K \) is written uniquely as \( a + b\alpha \), with \( a, b \in \mathbb{F}_2 \) (we identify any \( a \in \mathbb{F}_2 \) with its image \( a + \langle f \rangle \) in \( \mathbb{F}_2[X]/\langle f \rangle \)). Since \( \alpha \) is a root of \( f \), we have \( \alpha^2 = \alpha + 1 \). Thus, the addition and multiplication rules in \( K \) are given by the following rules:

\[
(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha,
\]

\[
(a + b\alpha) \cdot (c + d\alpha) = ac + (ad + bc)\alpha + bd\alpha^2 =
\]

\[
ac + (ad + bc)\alpha + bd(\alpha + 1) = (ac + bd) + (ad + bc + bd)\alpha,
\]

for any \( a, b \in \mathbb{F}_2 \).

Conversely, if \( L \) is a field with \( p^n \) elements and \( \alpha \in L \) is a generator of \( (L^*, \cdot) \) (see 3.4), then \( \text{Irr}(\alpha, \mathbb{F}_p) \) is an irreducible polynomial of degree \( n \) in \( \mathbb{F}_p[X] \).

**1.25 Corollary.** Let \( K \) be a field and let \( f \in K[X] \), \( \deg f = n \geq 1 \). Then there exists an extension \( L \) of \( K \) such that \( f \) is written as a product of polynomials of degree 1 in \( L[X] \) (one also says “\( f \) splits over \( L \)” or “\( f \) has \( n \) roots in \( L \)”).

**Proof.** Induction by the degree of \( f \). If \( \deg f = 1 \), take \( L = K \). Suppose that, for any field \( F \) and any \( g \in F[X] \) with \( \deg g < n \), there exists an extension of \( F \) in which \( g \) splits. Let \( f \in K[X] \), \( \deg f = n \). There ex-
ists an extension \( E \) of \( K \) in which \( f \) has a root \( a \). So \( X - a \) divides \( f \) in \( E[X] \), hence \( f = (X - a)g \), for some \( g \in E[X] \). Apply now the induction hypothesis for the field \( E \) and \( g \in L[X] \) to obtain an extension field \( L \) of \( E \) in which \( g \) splits. Obviously, \( f \) splits also in \( L \).  

For a given \( f \in K[X] \), the “minimal” extension of \( K \) over which \( f \) splits is called the \textit{splitting field} of \( f \) over \( K \). We will study this concept in IV.2.14.

We saw that “\( x \) is algebraic over \( K \)” is equivalent to “\( K(x) \) is a finite dimensional \( K \)-vector space”. This translates the property of an element of being algebraic into a Linear Algebra property, concerning the dimension of a vector space. The following theorem and its consequences illustrate the strength of this translation.

\[ 1.26 \text{ Theorem} \ (\text{The transitivity of finite extensions}) \ a) \text{ Let } K \subseteq L \text{ and } L \subseteq M \text{ be a tower of finite field extensions. Then } K \subseteq M \text{ is a finite extension and the degree is “multiplicative”:} \]

\[ [M : K] = [M : L][L : K] \]

Moreover, if \( \{x_1, \ldots, x_m\} \) is a \( K \)-basis of \( L \) and \( \{y_1, \ldots, y_n\} \) is a \( L \)-basis of \( M \), then \( \{x_1y_1, \ldots, x_1y_n, \ldots, x_my_1, \ldots, x_my_n\} \) is a \( K \)-basis of \( M \).

b) If \( K \subseteq M \) is a finite extension of fields and \( L \) is an intermediate field, then \( K \subseteq L \) and \( L \subseteq M \) are finite extensions.

\textbf{Proof.} a) It is sufficient to prove the claim on bases. Let us show that \( \{x_iy_j \mid 1 \leq i \leq m, 1 \leq j \leq n\} \) generates the \( K \)-vector space \( L \). If \( z \in M \), because \( \{y_1, \ldots, y_n\} \) is an \( L \)-basis, there exists \( b_1, \ldots, b_n \in L \) such that \( z = b_1y_1 + \ldots + b_ny_n \). Each \( b_j \ (1 \leq j \leq n) \) is of the form \( b_j = a_{j1}x_1 + \ldots + a_{jm}x_m \), for some \( a_{ji} \in K \ (1 \leq i \leq m) \). Using this in the expression for \( z \), we obtain that \( z \) is a linear combination of \( \{x_iy_j \mid 1 \leq i \leq m, 1 \leq j \leq n\} \) with coefficients in \( K \). Let us prove that \( \{x_iy_j \mid 1 \leq i \leq m, 1 \leq j \leq n\} \) is \( K \)-linearly independent.
If \( \sum_{1 \leq i \leq m, 1 \leq j \leq n} a_{ij} x_i y_j = 0 \), with \( a_{ij} \in K \), then \( \sum_{1 \leq i \leq m} \left( \sum_{1 \leq j \leq n} a_{ij} x_i \right) y_j = 0 \), where \( \sum_{1 \leq j \leq n} a_{ij} x_i \in L \). The \( L \)-linear independence of \( \{y_1, \ldots, y_n\} \) implies that \( \sum_{1 \leq j \leq n} a_{ij} x_i = 0 \), for any \( i \). Hence \( a_{ij} = 0 \) for any \( i \) and \( j \), because \( \{x_1, \ldots, x_m\} \) are \( K \)-linearly independent.

\[ \square \]

**1.27 Remarks.**

a) The theorem is also true for division rings, since the proof does not use the commutativity of multiplication.

b) More generally, if the extensions are not necessarily finite, suppose that \( (x_i)_{i \in I} \) is a \( K \)-basis of \( L \) and \( (y_j)_{j \in J} \) is an \( L \)-basis of \( M \). Then \( (x_i y_j)_{(i,j) \in I \times J} \) is a \( K \)-basis of \( M \).

**1.28 Examples.**

a) Consider \( \mathbb{Q}(\sqrt{2})/\mathbb{Q} \). \( \text{Irr}(\sqrt{2}, \mathbb{Q}) = X^2 - 2 \), using Eisenstein’s criterion (or observing that \( X^2 - 2 \) has no rational roots). Thus, the degree of the extension is 3 and a \( \mathbb{Q} \)-basis of \( \mathbb{Q}(\sqrt{2}) \) is \( \{1, \sqrt{2}, \sqrt{4}\} \). This means that any element of \( \mathbb{Q}(\sqrt{2}) \) is written uniquely as \( a + b \sqrt{2} + c \sqrt{4} \), with \( a, b, c \in \mathbb{Q} \). The same argument shows that for any \( n \in \mathbb{N}^* \), the extension \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \) has degree \( n \) and a \( \mathbb{Q} \)-basis is \( \{1, \sqrt{2}, \ldots, \sqrt{2^{n-1}}\} \).

b) The extension \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3}) \) has degree 4 and a \( \mathbb{Q} \)-basis is \( \{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\} \). For the proof, consider the tower of extensions \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2})(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3}) \). The first extension has degree 2 (since \( \text{Irr}(\sqrt{2}, \mathbb{Q}) = X^2 - 2 \) and a basis is \( \{1, \sqrt{2}\} \). The extension \( \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2})(\sqrt{3}) \) has also degree 2, because \( \text{Irr}(\sqrt{3}, \mathbb{Q}(\sqrt{2})) = X^2 - 3 \). Indeed, \( X^2 - 3 \) is irreducible over \( \mathbb{Q}(\sqrt{2}) \), because \( \sqrt{3} \notin \mathbb{Q}(\sqrt{2}) \): if, by absurd, \( \sqrt{3} = a + b \sqrt{2} \), with \( a, b \in \mathbb{Q} \), then \( 3 = a^2 + 2b^2 + 2ab \sqrt{2} \). If \( ab \neq 0 \), then \( \sqrt{2} \in \mathbb{Q} \), contradiction. If \( ab = 0 \), then \( 3 = a^2 \) or \( 3 = 2b^2 \), with \( a, b \in \mathbb{Q} \) (again a contradiction).
Thus, a basis of the extension $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2})(\sqrt{3})$ is $\{1, \sqrt{3} \}$. Applying the method in the theorem of transitivity of finite extensions, a basis of the extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6} \}$. Note that this extension has the proper intermediate fields $\mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{3}), \mathbb{Q}(\sqrt{6})$. These are all the proper intermediate fields. At this stage, this is difficult to prove, but it is an immediate consequence of Galois theory.

The element $\sqrt{2} + \sqrt{3} =: \alpha$ is a primitive element: $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. The inclusion $\mathbb{Q}(\sqrt{2} + \sqrt{3}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is evident. Let us show that $\sqrt{2}, \sqrt{3} \in \mathbb{Q}(\alpha)$. We have $\alpha^2 = 5 + 2\sqrt{6}$, so $\sqrt{6} \in \mathbb{Q}(\alpha)$. Thus $\sqrt{6}\alpha = 3\sqrt{2} + 2\sqrt{3} =: \beta \in \mathbb{Q}(\alpha)$. The system
\[
\begin{align*}
3\sqrt{2} + 2\sqrt{3} &= \beta \\
\sqrt{2} + \sqrt{3} &= \alpha
\end{align*}
\]
where $\alpha, \beta \in \mathbb{Q}(\alpha)$, shows that $\sqrt{2}, \sqrt{3} \in \mathbb{Q}(\alpha)$.

Let us find $\text{Irr}(\alpha, \mathbb{Q})$. The polynomial $\text{Irr}(\alpha, \mathbb{Q})$ has degree equal to
\[
[\mathbb{Q}(\alpha) : \mathbb{Q}] = [\mathbb{Q}(\alpha) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \cdot 2 = 4.
\]
We have $\alpha^2 = 5 + 2\sqrt{6} \Rightarrow (2\sqrt{6})^2 = (\alpha^2 - 5)^2 \Rightarrow (\alpha^2 - 5)^2 - 24 = 0$.
Thus, $\alpha$ is the root of $(X^2 - 5)^2 - 24$, which is monic, in $\mathbb{Q}[X]$ and has degree 4, so it must be equal to $\text{Irr}(\alpha, \mathbb{Q})$.

c) Here is a nontrivial example of a finite extension. Let $K$ be a field, $K(T)$ the field of rational fractions with coefficients in $K$, $u \in K(T)$, $u \not\in K$ and $f, g \in K[T]$ with $u = f/g$ and $(f, g) = 1$. Then the extension $K(u) \subseteq K(T)$ is finite and $[K(T) : K(u)] = \max(\deg f, \deg g)$.

Let $h = g(X)u - f(X) \in K(u)[X]$ (where $g(X)$ is the polynomial in $X$ obtained by replacing $T$ with $X$ in the polynomial $g$). We have $h(T) = 0$, which shows that $T$ is algebraic over $K(u)$, so $K(u) \subseteq K(T)$ is finite.

Since $K(T) = K(u)(T)$, $[K(T) : K(u)] = \deg \text{Irr}(T, K(u))$. We claim that $\text{Irr}(T, K(u))$ is associated to $h = g(X)u - f(X) \in K(u)[X]$. 

We have \( \deg h = \max(\deg f, \deg g) \). This is clear if \( \deg f \neq \deg g \); if \( \deg f = \deg g \), then the leading coefficient of \( h \) is \( ub - a \), where \( a \), respectively \( b \), are the leading coefficients of \( f \), respectively \( g \). We have \( ub - a \neq 0 \), because otherwise \( u = ab^{-1} \in K \), contradicting the assumption \( u \notin K \).

We have to prove that \( h = g(X)u - f(X) \in K(u)[X] \) is irreducible. First, note that \( u \) is transcendental over \( K \). Indeed, \( K \subseteq K(T) \) is infinite (\( T \) is transcendental over \( K \)) so \( K \subseteq K(u) \) must also be infinite (otherwise, considering the tower of extensions \( K \subseteq K(u) \subseteq K(T) \), the theorem of transitivity of finite extensions would imply that \( K \subseteq K(T) \) is finite). Thus, there is a \( K \)-isomorphism \( K[Y] \cong K[u] \) (where \( Y \) is an indeterminate). The irreducibility of \( g(X)u - f(X) \in K(u)[X] \) is thus equivalent to the irreducibility of \( r = g(X)Y - f(X) \in K(Y)[X] \). Since \( K[Y] \) is a UFD and \( K(Y) \) is its field of quotients, this is equivalent to \( r \) being irreducible in \( K[Y][X] \cong K[Y, X] \). Suppose \( r = pq \), with \( p, q \in K[Y, X] \). Let \( \deg_Y r \) be the degree of \( r \), seen as a polynomial in \( Y \) with coefficients in \( K[X] \). We have \( 1 = \deg_Y r = \deg_Y p + \deg_Y q \), so we may suppose \( \deg_Y p = 1 \) and \( \deg_Y q = 0 \). So, \( q \in K[X] \) and \( p = CY + d \), for some \( c, d \in K[X] \) and \( r = g(X)Y - f(X) = (CY + d)q \). Identifying the coefficients of the powers of \( Y \), we obtain that \( q|g \) and \( q|f \) in \( K[X] \), implying that \( q \in K^* \), since \( (f, g) = 1 \). Thus, \( h \) is irreducible in \( K[Y][X] \).

The polynomial \( h \) is irreducible and vanishes in \( T \), so it is associated to \( \text{Irr}(T, K(u)) \). Proposition 1.20 says that \([K(T) : K(u)] = \deg \text{Irr}(T, K(u)) = \deg h = \max(\deg f, \deg g)\).

It is interesting to remark that any intermediate field \( K \subseteq L \subseteq K(T) \) with \( K \neq L \) is of the form \( L = K(u) \), for some \( u \in K(T), u \notin K \). This fact is known as “Lüroth’s Theorem” and is of significance in Algebraic Geometry (see Morandi [1996], Walker [1950]).

1.29 Definition. A field extension \( K \subseteq L \) is called algebraic if each element in \( L \) is algebraic over \( K \). The extension \( K \subseteq L \) is
1.30 Proposition. Any finite extension is algebraic and finitely generated.

Proof. Let $K \subseteq L$ be finite and let $x \in L$. Then $K \subseteq K(x)$ is finite (as an intermediate field of $K \subseteq L$) and Prop. 1.20 shows that $x$ is algebraic over $K$. If \{$x_1, \ldots, x_m$\} is a $K$-basis of $L$, then certainly $L = K(x_1, \ldots, x_m)$, so $L$ is a finitely generated extension of $K$.

1.31 Proposition. Suppose $K \subseteq L$ is a field extension, $n \in \mathbb{N}^*$ and $x_1, \ldots, x_n \in L$ are algebraic over $K$. Then $K \subseteq K(x_1, \ldots, x_n)$ is a finite extension. In particular, any algebraic finitely generated extension is a finite extension.

Proof. We prove the claim by induction on $n \in \mathbb{N}^*$. For $n = 1$, the conclusion follows from Prop. 1.20. If $n > 1$, the induction hypothesis says that $K \subseteq K(x_1, \ldots, x_{n-1})$ is finite. Since $x_n$ is algebraic over $K$, $x_n$ is also algebraic over $K(x_1, \ldots, x_{n-1})$; the extension $K(x_1, \ldots, x_{n-1}) \subseteq K(x_1, \ldots, x_{n-1})(x_n)$ is thus finite. The conclusion follows by applying the transitivity of finite extensions to the tower $K \subseteq K(x_1, \ldots, x_{n-1}) \subseteq K(x_1, \ldots, x_{n-1}, x_n)$.

An easy argument by induction on $n \in \mathbb{N}^*$ shows that, if $K \subseteq L$ is an extension and $x_1, \ldots, x_n \in L$ are algebraic over $K$, then $K[x_1, \ldots, x_n] = K(x_1, \ldots, x_n)$ (see exercise 1.17). The converse of this result is nontrivial and it is known as “Zariski’s Lemma” (see for instance SPINDLER [1994], Prop. 12.35, p. 232).

1.32 Theorem. (transitivity of algebraic extensions) If $K \subseteq L$ and $L \subseteq M$ are algebraic extensions, then $K \subseteq M$ is an algebraic extension.
**Proof.** Let \( x \in M \). In order to prove that \( x \) is algebraic over \( K \), we show that \( x \) is contained in a finite extension of \( K \) and apply then Prop. 1.20.

Since \( x \) is algebraic over \( L \), there exists \( 0 \neq f \in L[X] \) such that \( f(x) = 0 \). Let \( a_0, \ldots, a_n \in L \) be the coefficients of \( f \). Then \( x \) is algebraic over \( K(a_0, \ldots, a_n) \) and \( K \subseteq K(a_0, \ldots, a_n) \) is finite, by Prop. 1.31. We have now the tower of finite extensions \( K \subseteq K(a_0, \ldots, a_n) \subseteq K(a_0, \ldots, a_n)(x) \). By transitivity, \( K \subseteq K(a_0, \ldots, a_n)(x) \) is also finite. \( \square \)

**1.33 Example.** The extension \( \mathbb{Q} \subseteq L = \mathbb{Q}\left(\frac{n}{\sqrt{2}} \middle| n \in \mathbb{N}^*\right) \) is algebraic and infinite. \( \mathbb{Q} \subseteq L \) is algebraic because \( L \) is the union of its intermediate fields of the form \( \mathbb{Q}\left(\frac{m}{\sqrt{2}} \middle| m \leq n\right) \), \( m \in \mathbb{N}^* \), and each of these is algebraic (they are even finite). But \( \mathbb{Q} \subseteq L \) is not finite: if, by contradiction, \( [L : \mathbb{Q}] = n \in \mathbb{N}^* \), then each intermediate field would have the degree a divisor of \( n \). But \( \mathbb{Q} \subseteq \mathbb{Q}\left(\frac{n+1}{\sqrt{2}}\right) \) has degree \( n + 1 \).

**1.34 Definition.** If \( K \subseteq L \) is a field extension, the set of all elements in \( L \) that are algebraic over \( K \) is denoted by \( K' \) and is called the algebraic closure of \( K \) in \( L \). Clearly, \( K \subseteq K' \).

**1.35 Proposition.** Let \( K \subseteq L \) be a field extension. Then \( K' \) is a subfield of \( L \) and an extension of \( K \). In particular, the sum, the difference, the product and the quotient (if it exists) of two algebraic elements over \( K \) are also algebraic elements over \( K \).

**Proof.** If \( a, b \in L \), with \( b \neq 0 \), are algebraic over \( K \), then \( K \subseteq K(a, b) \) is finite, so it is an algebraic extension. Thus \( a + b, a - b, ab, ab^{-1} \) are algebraic over \( K \), as elements of \( K(a, b) \). This shows that \( K' \) is a subfield of \( L \). \( \square \)

**1.36 Example.** \( \mathbb{Q}' \) is a subfield in \( \mathbb{C} \), an infinite algebraic extension of \( \mathbb{Q} \). Why?
Exercises

1. Let $K \subseteq L$ be an extension. Show that $L$ is a $K$-vector space with respect to: the addition of $L$ and the “scalar multiplication” given by: $\forall \alpha \in K, \forall x \in L, \alpha x = \alpha x$ (the multiplication in $L$). Prove that $[L : K] = 1$ if and only if $K = L$.

2. If $S$ is a subset of the field $L$ and $K$ is a subfield in $L$, then the field of fractions of the domain $K[S]$ is canonically isomorphic to $K(S)$.

3. Let $K \subseteq L$ be an extension. Consider the linear system with coefficients $a_{ij}, b_i \in K$:

\[
(S): \begin{cases}
a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\
a_{21}x_1 + \ldots + a_{2n}x_n = b_2 \\
\vdots \\
a_{m1}x_1 + \ldots + a_{mn}x_n = b_m
\end{cases}
\]

Prove that: the system $(S)$ has a solution in $E \iff (S)$ has a solution in $K$. Can you prove other similar properties of $(S)$?

4. Let $K \subseteq L$ be an extension and let $p \in L[X] \setminus K[X]$, deg $p = n$. Prove that $|p(K) \cap K| \leq n$. (Hint. Suppose $p(\alpha_i) = \beta_i, 1 \leq i \leq n + 1$, with $\alpha_i, \beta_i \in K$. Interpret this as a system of $n + 1$ equations, the unknowns being the coefficients of the polynomial.)

5. Let $K \subseteq L$ be an extension and let $\alpha \in L$, algebraic over $K$. If deg Irr$(\alpha, K) = n$, then deg Irr$(\beta, K)$ divides $n$, for any $\beta \in K(\alpha)$.

6. Let $K \subseteq L$ be an extension and let $x \in L$. Prove that $x$ is transcendental over $K$ if and only if $K[x]$ is $K$-isomorphic to the $K$-algebra $K[X]$ of polynomials in the indeterminate $X$.

7. Let $K$ be a field and let $(K_i)_{i \in I}$ be a chain of subfields of $K$ ($\forall i, j \in I, K_i \subseteq K_j$ or $K_j \subseteq K_i$). Prove that $\bigcup_{i \in I} K_i$ is a subfield in $K$.

8. Find a basis of the extension $\mathbb{Q} \subseteq \mathbb{Q} \left[ \sqrt[3]{3} \right]$. Express $(1 + \sqrt[3]{3})^{-1}$ in this basis. Find Irr$(1 + \sqrt[3]{3}, \mathbb{Q})$. The same problem for $\mathbb{Q} \subseteq \mathbb{Q}(\alpha)$ and the element $1 - \alpha$, where $\alpha$ is a root of $X^3 + X + 1$. 
9. Let $K \subseteq K(\alpha)$ be an algebraic extension. Suppose $\text{Irr}(\alpha, K) = f$ is known, where $\deg f = n$. For $0 \neq \beta \in K(\alpha)$ given as a $K$-linear combination of $1, \alpha, \ldots, \alpha^{n-1}$, describe a method to find $\beta^{-1}$ and $\text{Irr}(\beta, K)$.

10. Let $K \subseteq L$ be an extension of degree 2 ("quadratic extension"). Show that $L = K(\alpha)$, where $\alpha \in L$ is a root of a polynomial of the form $X^2 - b$ (if $\text{char } K \neq 2$), respectively $X^2 - b$ or $X^2 - X - b$ (if $\text{char } K = 2$), for some $b \in K$. Prove that any quadratic extensions of $\mathbb{Q}$ is of the form $\mathbb{Q}(\sqrt{d})$, with $d \in \mathbb{Z}$ squarefree ($d \neq 0, d \neq 1$ and $d$ is not the multiple of any square of a prime).

11. Let $m, n \in \mathbb{Z}$ be distinct and squarefree. Then $\mathbb{Q}(\sqrt{m}, \sqrt{n})$ has degree 4 over $\mathbb{Q}$. (Hint. Show that $\sqrt{m} \notin \mathbb{Q}(\sqrt{n})$.)

12. Let $K \subseteq L$ be an extension such that $L = K(x, y)$, with $x, y \in L$. If there exists $m, n \in \mathbb{N}^*$ such that $x^m \in K$, $y^n \in K$ and $(m, n) = 1$, then $K(x, y) = K(xy)$.

13. Let $K$ and $L$ be fields and let $\varphi : K \rightarrow L$ be a field isomorphism. Let $\psi : K[X] \rightarrow L[X]$ be the ring isomorphism that extends $\varphi$ and with $\psi(X) = X$. Prove that $p \in K[X]$ has a root in $K \iff \psi(p)$ has a root in $L$.

14. Let $K$ and $L$ be extensions of $\mathbb{Q}$ and let $\varphi : K \rightarrow L$ be a field homomorphism. Show that $\varphi|_{\mathbb{Q}} = \text{id}$. Prove a similar property for extensions of $\mathbb{F}_p$ (where $p$ is a prime).

15. Let $d, e \in \mathbb{Z}$, squarefree. Prove that $\mathbb{Q}(\sqrt{d}) \cong \mathbb{Q}(\sqrt{e}) \iff d = e$.

16. Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a field homomorphism.
   a) Prove that $\forall x \in \mathbb{R}, \ x > 0$ implies $\varphi(x) > 0$. Deduce $\varphi$ is an increasing map. (Hint. $x = a^2$, for some $a \in \mathbb{R}$.)
   b) Prove that $\varphi = \text{id}$. (Hint. Use that $\varphi|_{\mathbb{Q}} = \text{id}$ and that between any two real numbers there is a rational number.)

17. Let $K \subseteq L$ be an extension, $n \in \mathbb{N}^*$ and let $x_1, \ldots, x_n \in L$, algebraic over $K$. Then $K[x_1, \ldots, x_n] = K(x_1, \ldots, x_n)$. More generally, if $S$ is a subset of $L$ consisting of algebraic elements over $K$, then $K[S] = K(S)$. 
18. Let $E$ and $F$ be intermediate fields of the extension $K \subseteq L$. Prove that $E(F)$ ( = the composite $EF$) is equal to the set $\{xy^{-1} \mid x, y \text{ linear combinations of elements in } F \text{ with coefficients in } E, y \neq 0\}$. Under what assumptions is the composite $EF$ equal to the set of all linear combinations of elements in $F$ with coefficients in $E$?

19. Let $L/K$ be an extension. Show that $\text{IF}(L/K) = \{E \mid E \text{ is a subfield in } L, K \subseteq E\}$ is a complete lattice with respect to inclusion (any family of subfields $(E_i)_{i \in I}$ has sup and inf in $\text{IF}(L/K)$). What is the form of an arbitrary element in sup$\{E_i \mid i \in I\}$?

20. Let $K \subseteq L$ be an algebraic extension. Show that any subring of $L$ that includes $K$ (any $K$-subalgebra of $L$) is a field. Is the converse true?

21. Let $R \subseteq S$ be an extension of domains ($R$ is a subring of the domain $S$ and $1 \in R$). An $x \in S$ is called integral over $R$ if $x$ is the root of a monic polynomial in $R[X]$. Let $x \in \mathbb{C}$ be integral over $\mathbb{Z}$. Prove that $\text{Irr}(x, \mathbb{Q}) \in \mathbb{Z}[X]$. Deduce that the integral elements in $\mathbb{Q}(\sqrt{d})$, where $d \in \mathbb{Z}$ is squarefree, are roots of a monic polynomial of degree 2 with integer coefficients.

22. Let $K \subseteq L$ be a field extension and let $x, y \in L$ be algebraic over $K$. Let $p = \text{Irr}(x, K)$, $q = \text{Irr}(y, K)$, deg $p = m$, deg $q = n$.
   a) Prove that $\deg \text{Irr}(x + y, K) \leq mn$ and $\deg \text{Irr}(xy, K) \leq mn$.
   b) $\deg \text{Irr}(x, K(y)) = m \iff \deg \text{Irr}(y, K(x)) = n \iff [K(x, y) : K] = mn$.
   c) Formulate a sufficient condition for $[K(x, y) : K] = mn$.

23. Find $\text{Irr}(x, \mathbb{Q})$ and a basis of the extension $\mathbb{Q} \subseteq \mathbb{Q}(x)$, $\forall x \in \{\sqrt{3} - i, \sqrt{1 + \sqrt{3}}, (1 - i)\sqrt{2}\}$.

24. For any $n \in \mathbb{N}^*$, give an example of extension of degree $n$ of $\mathbb{Q}$.

25. Let $K \subseteq L \subseteq E$ be a tower of extensions and let $x \in E$ be algebraic over $K$. Then $[L(x) : L] \leq [K(x) : K]$.

26. Let $L$ and $E$ be intermediate fields of the extension $K \subseteq F$. 
IV.2 Roots of polynomials. Algebraically closed fields

a) If $K \subseteq L$ is algebraic, then $E \subseteq EL$ is algebraic.

b) If $S$ is a subset of $L$ and $K < S > = L$ ($S$ generates the $K$-vector space $L$), then $E < S > = EL$. (Hint. Reduce to the case when $S$ is finite.)

c) If $[L : K]$ is finite, then $[EL : E] \leq [L : K]$.

d) If the degrees $[L : K]$ and $[E : K]$ are finite and coprime, then $[EL : E] = [L : K]$ and $[EL : K] = [E : K] \cdot [L : K]$.

e) If $[EL : K] = [E : K] \cdot [L : K]$, then $K = L \cap E$.

f) If $[E : K] = 2$ and $K = L \cap E$, then $[EL : K] = [E : K] \cdot [L : K]$.

g) Give an example such that $[E : K] = [L : K] = 3$, $K = L \cap E$, but $[EL : K] < 9$.

27. Let $K \subseteq L$ be an extension of degree $n$ and let $g \in K[X]$, deg $g = p$, with $p$ prime, $(p, n) = 1$. If $g$ is irreducible in $K[X]$, then $g$ is irreducible in $L[X]$.

28. Show that $X^5 - 2$ is irreducible in $\mathbb{Q}(\omega)[X]$, where $\omega = \cos(2 \pi/5) + i \cdot \sin(2 \pi/5)$.

IV.2 Roots of polynomials. Algebraically closed fields

Let $R$ be a domain. Bézout's theorem says that the polynomial $f \in R[X]$ has the root $a \in R$ if and only if $X - a$ divides $f$ in $R[X]$. It is thus natural to consider the following definition:
2.1 Definition. Suppose $R$ is a domain, $f \in R[X]$ is a nonzero polynomial, $a \in R$ and $n \in \mathbb{N}$. We say that $a$ is a multiple root of $f$ with multiplicity $n$ if $(X - a)^n | f$ and $(X - a)^{n+1} \nmid f$. The natural number $n$ is called the multiplicity of the root $a$. If $n = 1$, $a$ is called a simple root of $f$. If $n > 1$, $a$ is called a multiple root of $f$ (double if $n = 2$, triple if $n = 3$ etc).

When counting the roots of a polynomial, each root counts with its multiplicity (unless otherwise specified).

2.2 Proposition. Let $R$ be a UFD and let $f \in R[X]$ be nonzero. If $a_1, \ldots, a_n \in R$ are distinct roots of $f$, of multiplicities $m_1, \ldots, m_n$ respectively, then $(X - a_1)^{m_1} \cdots (X - a_n)^{m_n}$ divides $f$ in $R[X]$.

Proof. Use induction on $n$. If $n = 1$, this is the definition. Suppose the claim is true for $n - 1$ and let $f$ be as in the statement. By the induction hypothesis, $f = (X_1 - a_1)^{m_1} \cdots (X_{n-1} - a_{n-1})^{m_{n-1}} g$, for some $g \in R[X]$. Since the polynomials $X - a_i$, $1 \leq i \leq n$, are irreducible and pairwise not associated in divisibility, $(X - a_1)^{m_1}$, ..., $(X - a_n)^{m_n}$ are pairwise relatively prime. By Prop. I.5.6, $(X - a_n)^{m_n}$ is coprime with $(X - a_1)^{m_1} \cdots (X - a_{n-1})^{m_{n-1}}$. Since $(X - a_n)^{m_n}$ divides $(X - a_1)^{m_1} \cdots (X - a_{n-1})^{m_{n-1}} g$, we obtain that $(X - a_n)^{m_n}$ divides $g$.

2.3 Corollary. Let $R$ be a domain and let $f \in R[X]$, $\deg f = n$. Then $f$ has at most $n$ roots in $R$.

Proof. Of course, each root is counted with its multiplicity. Let $K$ be the field of fractions of $R$. We see $f$ as a polynomial in $K[X]$ and apply the preceding proposition.

We now seek a criterion to decide if a polynomial has multiple roots. It is useful to introduce the notion of formal derivative of a polynomial with coefficients in an arbitrary ring.
2.4 Definition. Let $R$ be a commutative unitary ring and let 
$$f = a_0 + a_1X + \ldots + a_nX^n \in R[X].$$

The (formal) derivative of $f$ is the polynomial 
$$df := a_1 + 2a_2X + \ldots + na_nX^{n-1}.$$ 

Other notations: $df = f'$ or $df = f^{(1)}$.

Direct calculations show that the formal derivative has the usual properties of the derivative of a function as known from calculus:
$$\begin{align*}
(f + g)' &= f' + g', \\
(af)' &= af', \\
(fg)' &= f'g + fg',
\end{align*}$$
for all $a \in R$, $f, g \in R[X]$.

Note that $d : R[X] \to R[X]$ is the unique $R$-module homomorphism with the property that $d1 = 0$ and $dX^n = nX^{n-1}$, $\forall n \in \mathbb{N}^*$ (apply the universality property of a free module). Evidently, $\deg df \leq \deg f - 1$.

Composing the homomorphism $d$ with itself $n$ times ($n \in \mathbb{N}^*$) is denoted by $d^n$; $d^n : R[X] \to R[X]$. Thus, $d^n = d \circ d^{n-1}$, $\forall n \in \mathbb{N}^*$, $d^0 = \text{id}$.

$d^n f$ is often denoted by $f^{(n)}$, $\forall f \in R[X]$.

2.5 Proposition. Let $R$ be a domain, let $f \in R[X]$ of degree $n > 0$ and let $\alpha \in R$.

a) There exist and are unique $b_0, \ldots, b_n \in R$ such that 
$$f = \sum b_i (X - \alpha)^i.$$ 

b) If $\alpha$ is a root of multiplicity $m$ ($m \in \mathbb{N}^*$) of $f$, then $f^{(i)}(\alpha) = 0$, for any $i \in \{0, \ldots, m - 1\}$.

c) If $f(\alpha) = f'(\alpha) = 0$, then $\alpha$ is a multiple root of $f$ (of multiplicity at least 2).

Proof. a) Use induction on $\deg f$. If $f = a_0 + a_1X$, then $f = (a_0 + a_1\alpha) + a_1(X - \alpha)$. If $\deg f = n > 1$, by the integer division with remainder theorem applied to $f$ and $X - \alpha$ (I.6.2), we obtain $f = (X - \alpha)g + b_0$, with $b_0 \in R$ and $g \in R[X]$, $\deg g = n - 1$. Writing $g$ in the form given by the induction hypothesis and replacing in the previous equality, we obtain the result.

The uniqueness part amounts to the $R$-linear independence of $\{(X - \alpha)^i \mid i \in \mathbb{N}\}$ in $R[X]$, easy to prove.
b) By a), \((X - \alpha)^m\) divides \(f\) if and only if \(b_0, b_1, \ldots, b_{m-1}\) are all 0. On the other hand, \(f^{(i)}(\alpha) = i!b_i, \forall i \in \{0, \ldots, n\}\) (proof by induction on \(n\)). This implies that \(f^{(i)}(\alpha) = 0, \forall i \in \{0, \ldots, m-1\}\).

c) Let \(f\) be written as in a). We obtain \(f(\alpha) = b_0 = 0\) and \(f'(\alpha) = b_1 = 0\). So, \((X - \alpha)^2 \mid f\).

Suppose \(K\) is a field, \(\Omega\) is an extension of \(K\), \(\alpha \in \Omega\) and \(f \in K[X]\). Applying the result above, we obtain that \(\alpha\) is a multiple root of \(f\) if and only if is simultaneously a root of \(f\) and of its derivative: \((X - \alpha)|f\) and \((X - \alpha)|f'\) in \(\Omega[X]\). This implies that the GCD of \(f\) and \(f'\) in \(\Omega[X]\) has degree \(\geq 1\). But the GCD of two polynomials can be obtained using Euclid's algorithm and does not depend on the field considered: if \(f, g \in K[X]\), then \((f, g)_{K[X]} = (f, g)_{\Omega[X]}\). Thus:

2.6 Proposition. Let \(K\) be a field and let \(f \in K[X]\). Then \(f\) has multiple roots (in some extension of \(K\)) if and only if \(f\) and \(f'\) are not relatively prime.

Proof. Let \(\Omega\) be an extension of \(K\) in which \(f\) splits (IV.1.25). If \(f\) has multiple roots, we saw that \(f\) and \(f'\) are not relatively prime. Conversely, if \(g = \text{GCD}(f, f')\) has degree \(\geq 1\), then \(g\) has a root \(\alpha\) in \(\Omega\) (the roots of \(g\) are among the roots of \(f\)) and \(\alpha\) is a multiple root of \(f\) since \(f(\alpha) = f'(\alpha) = 0\).

Without knowing the roots of a polynomial, the criterion above allows to decide if it has multiple roots.

2.7 Proposition. (Viète's relations)\(^{44}\) Suppose \(R\) is a domain, \(f = a_0 + a_1X + \ldots + a_nX^n\) is a polynomial in \(R[X]\), \(a_n \neq 0\), having the roots \(x_1, \ldots, x_n \in R\). Then:

\[
f = a_n(X - x_1)\ldots(X - x_n)
\]

For any \(k, 1 \leq k \leq n\), we have:

\(^{44}\) In honor of François Viète, 1540-1603, French mathematician.
IV.2 Roots of polynomials. Algebraically closed fields

\[ a_n \sum_{|i_1, \ldots, i_k| \leq \{1, 2, \ldots, n\}} x_{i_1} \cdots x_{i_k} = (-1)^k a_{n-k}. \]

In particular,

\[ a_n (x_1 + \ldots + x_n) = -a_{n-1} \]
\[ a_n (x_1 x_2 + x_1 x_3 + \ldots + x_{n-1} x_n) = a_{n-2} \]
\[ \ldots \]
\[ a_n x_1 \ldots x_n = (-1)^n a_0. \]

Consequently, any root in \( R \) of \( f \) divides \( a_0 \).

**Proof.** The polynomial \( g = a_n (X - x_1) \ldots (X - x_n) \) divides \( f \), by prop. 2.2. The polynomials \( g \) and \( f \) have the same degree and \( g | f \), so they are associated in divisibility in \( K[X] \) (\( K \) is the field of fractions of \( R \)). Since \( g \) and \( f \) have the same leading coefficient, \( g = f \). The other equalities result by identifying the coefficients of \( g \) and \( f \). \( \square \)

2.8 Corollary. Let \( R \) be a subring of the domain \( S \) and let \( f \in R[X] \) be monic. If \( \deg f = n \) and \( f \) has the roots \( x_1, \ldots, x_n \in S \), then \( g(x_1, \ldots, x_n) \in R \), for any symmetric polynomial \( g \in R[X_1, \ldots, X_n] \).

**Proof.** By the fundamental theorem of the symmetric polynomials, there exists \( h \in R[X_1, \ldots, X_n] \) such that \( g = h(s_1, \ldots, s_n) \), where \( s_1, \ldots, s_n \) are the fundamental symmetric polynomials in the indeterminates \( X_1, \ldots, X_n \). The relations between roots and coefficients show that \( s_i(x_1, \ldots, x_n) \in R \), for any \( 1 \leq i \leq n \). Thus:

\[ g(x_1, \ldots, x_n) = h(s_1(x_1, \ldots, x_n), \ldots, s_n(x_1, \ldots, x_n)), \]

which is an element of \( R \). \( \square \)

The fields that have no proper algebraic extensions are highly interesting. They are characterized by the next theorem.

2.9 Theorem. Let \( K \) be a field. The following statements are equivalent:

a) There exist no proper algebraic extensions of \( K \).

b) There exist no proper finite extensions of \( K \).
c) For any extension $L$ of $K$, the algebraic closure of $K$ in $L$ coincides with $K$. ("$K$ is algebraically closed in $L$”).

d) Any polynomial of degree at least 1 with coefficients in $K$ has a root in $K$.

e) Any polynomial $f$ of degree $n \geq 1$ with coefficients in $K$ splits over $K$ ($f$ has $n$ roots in $K$).

f) The irreducible polynomials in $K[X]$ are the polynomials of degree 1.

Proof. $a) \Rightarrow b)$ Clear, since any finite extension is algebraic.

$b) \Rightarrow c)$ Let $x \in K'$. Then $K \subseteq K(x)$ is finite, so $K = K(x)$ and $x \in K$.

c) $\Rightarrow d)$ Let $f \in K[X]$, $\deg f \geq 1$. Prop. 1.23 ensures the existence of an extension $L$ of $K$ in which $f$ has a root $x$. Since $x$ is in $K'$, $x \in K$.

d) $\Rightarrow e)$ Suppose by contradiction that there exists $f \in K[X]$, $\deg f \geq 1$, such that $f$ does not split in $K[X]$. Choose $f$ of minimum degree with this property. By hypothesis, $f$ has a root $a \in K$, so $f = (X - a)g$, with $g \in K[X]$. But $\deg g < \deg f$, so $g$ splits over $K$. Since $f = (X - a)g$, $f$ also splits over $K$, contradiction.

e) $\Rightarrow f)$ Evident.

$f) \Rightarrow a)$ Let $L$ be an algebraic extension of $K$ and let $a \in L$. Then $\text{Irr}(a, K)$ is an irreducible polynomial in $K[X]$, hence it has degree 1, so $a \in K$.

2.10 Definition. A field satisfying the equivalent properties above is called an algebraically closed field. An extension $L$ of the field $K$ is called an algebraic closure of $K$ if $L$ is algebraically closed and $K \subseteq L$ is an algebraic extension.

2.11 Examples. a) A finite field is never algebraically closed. Indeed, if $F$ is a field with $n$ elements ($n \geq 2$), the polynomial $f = 1 + \prod_{a \in F} (X - a) \in F[X]$ has degree $n$ and $f(a) = 1, \forall a \in F$, so $f$ has no roots in $F$. 
b) The complex field $\mathbb{C}$ is algebraically closed. This fact is known as “The Fundamental Theorem of Algebra” and will be proven later. $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$, because $\mathbb{C}$ is uncountable and any algebraic extension of $\mathbb{Q}$ is countable (see next lemma). But $\mathbb{Q}'_C$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, is an algebraic closure of $\mathbb{Q}$. More generally, if $K \subseteq L$ is an extension and $L$ is algebraically closed, then $K'L$ is an algebraic closure of $K$ (prove this!).

c) If $K$ and $L$ are isomorphic fields and $K$ is algebraically closed, then $L$ is algebraically closed.

An important problem is the existence (and uniqueness) of an algebraic closure of a given field $K$. First, let us prove a technical lemma, also interesting in its own right.

2.12 Lemma. Let $K \subseteq L$ be an algebraic extension. Then $|L| \leq \max(|K|, |\mathbb{N}|)$.

Proof. Let $P$ be the set of all irreducible monic polynomials in $K[X]$ and let $a \in L$. Any $f \in P$ has a finite number $n(f)$ of roots in $L$. We establish an indexing of these, $\{a_1, \ldots, a_{n(f)}\}$. Define $\varphi : L \to P \times \mathbb{N}$ as follows: if $a \in L$, let $\varphi(a) = (f, r)$, where $f = \text{Irr}(a, K)$ and $a = a_r$.

The function $\varphi$ is injective, so $|L| \leq |P| \times |\mathbb{N}| = \max(|P|, |\mathbb{N}|)$. We have to show $|P| \leq \max(|K|, |\mathbb{N}|)$. Note that $P$ is the union of the disjoint sets $(P_n)_{n \geq 1}$, where $P_n$ is the set of polynomials in $P$ of degree $n$, so: $|P| = \bigcup_{n \geq 1} P_n = \sum_{n \geq 1} |P_n|$. Of course, $|P_n| \leq |K|^n$, since $\psi : P_n \to K^n$, $\psi(a_0 + a_1X + \ldots + a_{n-1}X^{n-1} + X^n) = (a_0, \ldots, a_{n-1})$, is injective. On the other hand, $|K|^n = |K|$ if $K$ is infinite, so $|P| = \sum_{n \geq 1} |P_n| \leq \sum_{n \geq 1} |K| = \ldots$.

---

45 This name (kept for historic reasons) expresses the concept that Algebra studies mainly complex numbers. This is no longer true since the 19th century, although complex numbers continue to play an important role in mathematics.
\(|K \times \mathbb{N}| = |K|\). If \(K\) is finite, then \(|K|^n\) is finite and \(|P| \leq \sum_{n \geq 1} |K|^n = |\mathbb{N}|\) (as a countable union of finite sets).

An immediate consequence is that any algebraic extension of \(\mathbb{Q}\) is countable; thus, the extension \(\mathbb{Q} \subseteq \mathbb{R}\) is not algebraic, because \(\mathbb{R}\) is not countable. This means that there exist real transcendental numbers (even uncountably many!).

2.13 Theorem. Let \(K\) be a field. Then an algebraic closure of \(K\) exists.

Proof. Th. 2.9.a) suggests looking for the algebraic closure as a maximal element\(^{46}\) in the “set” of all algebraic extensions of \(K\). The problem is that the algebraic extensions of \(K\) do not form a set! Fortunately, we can include \(K\) in a “sufficiently large” set \(M\) (that contains “all” algebraic extensions of \(K\)) and confine our search to algebraic extensions of \(K\) included in \(M\).

Let \(M\) be a set such that\(^{47}\) \(|M| > \max(|K|, |\mathbb{N}|)\) and \(K \subseteq M\).\(^{48}\) Let \(A\) be the set of algebraic extensions of \(K\) included in \(M\). The set \(A\) is ordered: for any \(F, L \in A\), \(F \leq L\) if and only if \(F\) is a subfield of \(L\). The set \(A\) is nonempty (\(K \in A\)) and is inductively ordered. Indeed, if \(B\) is a chain in \(A\), the union of the elements in \(B\) is in \(A\) (prove!) and is an upper bound for \(B\). Zorn's Lemma assures the existence of a maximal element \(F\) of \(A\).

We prove that \(F\) is an algebraic closure of \(K\). Let \(E\) be an algebraic extension of \(F\). By the transitivity property of algebraic extensions, \(E\) is an algebraic extension of \(K\). The previous lemma says that

\(^{46}\) You guessed, Zorn’s Lemma will be used. All proofs of the existence of the algebraic closure use some form of the Axiom of Choice.

\(^{47}\) This choice for \(|M|\) is suggested by the previous lemma. We want to be sure that \(M\) includes a "copy" of any algebraic extension of \(K\).

\(^{48}\) In fact, we need just an injective function \(\alpha : K \to M\), but we assume \(K \subseteq M\) to simplify notations.
IV.2 Roots of polynomials. Algebraically closed fields

We can then define \( \varphi : E \rightarrow M \), injective, such that \( \varphi|_F = \text{id}_F \). By transport of structure, \( \varphi(E) \) becomes a field (for instance, the addition in \( \varphi(E) \) is defined by \( \varphi(x) + \varphi(y) := \varphi(x + y) \), \( \forall x, y \in E \)). Then \( \varphi : E \rightarrow \varphi(E) \) is an \( F \)-isomorphism of fields and thus \( \varphi(E) \) is an algebraic extension of \( K \). So, \( \varphi(E) \in A \) and \( F \subseteq \varphi(E) \). Because \( F \) is maximal in \( A \), \( F = \varphi(E) \), so \( F = E \). This shows that \( F \) is algebraically closed.

For a field \( K \) and a polynomial (or a family of polynomials), it is interesting to study the “smallest” field over which the polynomial (respectively, any polynomial in the family) splits.

2.14 Definition. Let \( K \) be a field, let \( \Omega \) be an algebraic closure of \( K \) and let \( F \) be a family of polynomials in \( K[X] \). The splitting field of the family \( F \) over \( K \) is the field \( S \) obtained by adjoining to \( K \) all the roots in \( \Omega \) of the polynomials in \( F \),

\[
S := K(\{x \in \Omega \mid \exists f \in F \text{ such that } f(x) = 0\}).
\]

If \( F = \{f\} \), \( S \) is called the splitting field of \( f \) over \( K \). If \( f \in K[X] \) has the roots \( x_1, \ldots, x_n \in \Omega \), then the splitting field of \( f \) over \( K \) is \( K(x_1, \ldots, x_n) = K[x_1, \ldots, x_n] \).

Of course, every polynomial in the family \( F \) splits in the splitting field of \( F \) over \( K \).

2.15 Remarks. a) At first glance, the splitting field of a family \( F \) over \( K \) depends on the choice of an algebraic closure \( \Omega \) of \( K \). We shall prove though that any two splitting fields of \( F \) over \( K \) are \( K \)-isomorphic. This is the reason we say the (and not a) splitting field of \( F \) over \( K \).

b) The algebraic closure of \( K \) is the splitting field over \( K \) of the family of all nonconstant polynomials in \( K[X] \).

2.16 Examples. a) The splitting field of \( X^2 - 2 \) over \( \mathbb{Q} \) is \( \mathbb{Q}(\sqrt{2}) \).

b) The splitting field of \( X^2 - 2 \) over \( \mathbb{R} \) is \( \mathbb{R} \).
c) The splitting field of $X^3 - 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt[3]{2}, \omega)$, where $\omega \in \mathbb{C}$ is a root of $X^2 + X + 1$. Indeed, the roots of $X^3 - 2$ are $\sqrt[3]{2}, \omega \sqrt[3]{2}, \omega^2 \sqrt[3]{2}$ and $\mathbb{Q}(\sqrt[3]{2}, \omega) = \mathbb{Q}(\sqrt[3]{2}, \omega^2 \sqrt[3]{2})$.

d) If $f \in K[X]$ is a polynomial of degree $n$, and $L$ is its splitting field over $K$, then $[L : K] \leq n$.

Indeed, if $x_1, \ldots, x_n \in L$ are the roots of $f$, then $[K(x_1) : K] = \deg(\text{Irr}(x_1, K)) \leq \deg f = n$. Note that $L$ is a splitting field over $K(x_1)$ of $g := f/(X - x_1) \in K(x_1)[X]$. Since $\deg g = n - 1$, apply an induction to obtain that $[L : K(x_1)] \leq (n - 1)!$ and so $[L : K] \leq n!$.

Proving the uniqueness (up to a $K$-isomorphism) of the splitting field of a family of polynomials over $K$ requires some results on the extension of field homomorphisms. These results have also other important applications.

We shall use frequently the following elementary fact: if $\sigma: K \to L$ is a field homomorphism, then $\sigma$ has a unique extension to a ring homomorphism $\tau: K[X] \to L[X]$, namely $\tau(a_0 + a_1X + \ldots + a_nX^n) = \sigma(a_0) + \sigma(a_1)X + \ldots + \sigma(a_n)X^n$. This is the unique ring homomorphism $\tau: K[X] \to L[X]$ satisfying $\tau|_K = \sigma$ and $\tau(X) = X$. The existence and uniqueness of $\tau$ are a consequence of the universality property of the polynomial ring $K[X]$. By a harmless abuse of notation, the extension to $K[X]$ of the homomorphism $\sigma$ is denoted also by $\sigma$.

The following property is very simple, but has deep implications; in particular, it is instrumental in the determination of the Galois group of an extension.

2.17 Proposition. If $E$ and $L$ are extensions of $K$, $\alpha \in E$ is a root of $f \in K[X]$, and $\varphi: E \to L$ is a $K$-homomorphism, then $\varphi(\alpha)$ is also a root of $f$.

Proof. Let $f = a_0 + a_1X + \ldots + a_nX^n \in K[X]$. Then $f(\alpha) = a_0 + a_1\alpha + \ldots + a_n\alpha^n = 0$. Apply the $K$-homomorphism $\varphi$ to this equality:
IV.2 Roots of polynomials. Algebraically closed fields

\[ 0 = \varphi(a_0 + a_1 \alpha + \ldots + a_n \alpha^n) = a_0 + a_1 \varphi(\alpha) + \ldots + a_n \varphi(\alpha)^n = f(\varphi(\alpha)). \quad \square \]

The proposition below uses the isomorphism \( K[X]/(f) \cong K(\alpha) \), where \( \alpha \) is algebraic over \( K \) and \( f = \text{Irr}(\alpha, K) \). In other words, for an irreducible polynomial \( f \in K[X] \), regardless of the choice of an extension \( E \) in which \( f \) has a root and regardless of the choice of a root \( \alpha \in E \), \( K(\alpha) \) is the same (up to a \( K \)-isomorphism).

\[ \text{2.18 Proposition.} \quad \text{Suppose } \sigma: K \to K' \text{ is a field isomorphism, } f \in K[X] \text{ is irreducible and } \alpha \text{ is a root of } f \text{ in some extension } E \text{ of } K. \text{ If } \alpha' \text{ is a root of } f' := \sigma(f) \in K[X] \text{ in some extension } E' \text{ of } K', \text{ then } K(\alpha) \text{ is isomorphic to } K'(\alpha') \text{ by an isomorphism } \sigma' \text{ that extends } \sigma \text{ and } \sigma'(\alpha) = \alpha'. \]

\[ \text{Proof.} \quad \text{We saw that there exists a } K\text{-isomorphism } \eta: K[X]/(f) \to K(\alpha) \text{ such that } \eta(X + (f)) = \alpha. \text{ Similarly, there exists a } K'\text{-isomorphism } \gamma: K'[X]/(f') \to K'(\alpha'), \text{ such that } \gamma(X + (f')) = \alpha'. \text{ We also have an isomorphism } \varphi: K[X]/(f) \to K'[X]/(f'), \varphi(h + (f)) = \sigma(h) + (f'), \text{ for any } h + (f) \in K[X]/(f). \text{ Then } \gamma \circ \varphi \circ \eta^{-1} \text{ is the isomorphism } \sigma' \text{ we need:} \]

\[ K(\alpha) \xrightarrow{\eta^{-1}} K[X]/(f) \xrightarrow{\varphi} K'[X]/(f') \xrightarrow{\gamma} K(\alpha') \quad \square \]

The seemingly pointless degree of generality in the previous proposition (taking two isomorphic fields \( K \) and \( K' \) instead of just \( K \)) is in fact useful in the next Theorem.

\[ \text{2.19 Theorem.} \quad \text{Let } K \text{ be a field, let } L \text{ be an algebraic extension of } K, \text{ let } \Omega \text{ be an algebraic closure of } K \text{ and let } \sigma: L \to \Omega \text{ be a } K\text{-homomorphism. If } E \text{ is an algebraic extension of } L, \text{ then there exists a } K\text{-homomorphism } \tau: E \to \Omega \text{ that extends } \sigma. \]

\[ \text{Proof.} \quad \text{Let } E \text{ be the set of “extensions” of the homomorphism } \sigma, \]

\[ E = \{ (F, \varphi) \mid F \text{ is a subfield of } E, L \subseteq F, \varphi: F \to \Omega, \varphi \text{ is a homomorphism, } \varphi|_L = \sigma \}. \]
E is ordered by "\( \leq \)" defined by: \((F, \varphi) \leq (F', \varphi')\) if and only if \(F \subseteq F'\) and \(\varphi|_F = \varphi\). A straightforward proof shows that "\( \leq \)" is indeed an order relation. Moreover, if \(\{(F_i, \varphi_i)\}_{i \in I}\) is a chain in \(E\), it is bounded above by \((\bigcup_{i \in I} F_i, \varphi)\), where \(\varphi: \bigcup_{i \in I} F_i \to \Omega\) is defined by \(\varphi(x) = \varphi_i(x)\) if \(x \in F_i\) (this definition is independent of the choice of \(i \in I\) such that \(x \in F_i\)). Thus, \(E\) is inductively ordered and has, by Zorn's Lemma, a maximal element \((F, \varphi)\).

Let us prove that \(F = E\), which finishes the proof. If \(F \neq E\), pick \(x \in E \setminus F\). The element \(x\) is algebraic over \(F\); let \(f = \text{Irr}(x, F)\). Apply Prop. 2.18 to the following situation: \(\varphi: F \to \varphi(F)\) field isomorphism, \(f \in F[X]\), \(x\) is a root of \(f\) (in \(E\)), \(x'\) is a root of \(\varphi(f)\) (in \(\Omega\)). We obtain an isomorphism \(\varphi': F(x) \to \varphi(F)(x')\) that extends \(\varphi\). Since \(\varphi(F)(x') \subseteq \Omega\), \(\varphi'\) is an extension of \(\varphi\) to \(F(x)\), contradicting the maximality of \((F, \varphi)\).

2.20 Corollary. Let \(K\) be a field. Then:

a) Any two algebraic closures of \(K\) are \(K\)-isomorphic.

b) If \(f \in K[X]\), then any two splitting fields of \(f\) over \(K\) are \(K\)-isomorphic.

c) If \(F\) is a family of polynomials in \(K[X]\), then any two splitting fields of the family \(F\) over \(K\) are \(K\)-isomorphic.

d) Let \(\Omega\) be an algebraic closure of \(K\). Then any algebraic extension of \(K\) is \(K\)-isomorphic to a subfield of \(\Omega\) that includes \(K\).

Proof. a) Let \(\Omega\) and \(\Omega'\) be algebraic closures of \(K\). The canonical inclusion \(\iota: K \to \Omega\) has an extension to a \(K\)-homomorphism \(\sigma: \Omega \to \Omega'\) (by the previous theorem). The image of \(\sigma\) is a subfield \(\sigma(\Omega)\) of \(\Omega'\), isomorphic to \(\Omega\) (by \(\sigma\)), so it is algebraically closed. Since \(\Omega'\) is algebraic over \(\sigma(\Omega)\), \(\sigma(\Omega) = \Omega'\). Thus, \(\sigma\) is an isomorphism.

c) Let \(\Omega\) and \(\Omega'\) be algebraic closures of \(K\). Let \(R\) (respectively \(R'\)) the set of all roots in \(\Omega\) (respectively \(\Omega'\)) of the polynomials in \(F\). We have to prove that \(K(R)\) and \(K(R')\) are \(K\)-isomorphic. Let \(\sigma: \Omega \to \Omega'\) be a \(K\)-isomorphism, as in a). If \(f \in F\), and \(\alpha \in \Omega\) is a root of \(f\), then \(\sigma(\alpha)\) is a root of \(f\) in \(\Omega'\). This shows that \(\sigma(R) \subseteq R'\). Considering
\(\sigma^{-1}: \Omega' \to \Omega\), we obtain analogously that \(\sigma^{-1}(R') \subseteq R\), so \(\sigma\) establishes a bijection between \(R\) and \(R'\). If we remember the form of the elements in \(K(R)\), we obtain that \(\sigma(K(R)) = K(\sigma(R)) = K(R')\), so the restriction of \(\sigma\) to \(K(R)\) is a \(K\)-isomorphism between \(K(R)\) and \(K(R')\).

\(b)\) is a particular case of \(c)\).

\(d)\) Let \(K \subseteq L\) be algebraic. Then the canonical inclusion \(\iota: K \to \Omega\) extends to a \(K\)-homomorphism \(\varphi: L \to \Omega\) by 2.19. Thus, \(L\) is \(K\)-isomorphic to \(\varphi(L)\), a subfield of \(K\).

Part \(d)\) above says that a given algebraic closure \(\Omega\) of \(K\) includes “all” the algebraic extensions of \(K\).

2.21 Theorem. (the Fundamental Theorem of Algebra)\(^{49}\) The field \(\mathbb{C}\) is algebraically closed.

\textbf{Proof.} Let us prove first that any polynomial \(f\) of degree \(\geq 1\) with real coefficients has a complex root.

If \(\deg f\) is odd, then \(f\) has a real root. This follows from the well known fact from analysis: the polynomial function \(\varphi: \mathbb{R} \to \mathbb{R}\) associated to \(f\) has both positive and negative values (since the limits of \(\varphi\) at \(+\infty\) and \(-\infty\) are infinite and have opposite signs); the continuity of \(\varphi\) implies the existence of \(c \in \mathbb{R}\) with \(\varphi(c) = f(c) = 0\).

Let \(\deg f = n \in \mathbb{N}\) and let \(s(f)\) be the exponent of 2 in the prime factor decomposition of \(n\) (\(2^{s(f)} \mid n\) and \(2^{s(f) + 1} \nmid n\)). We prove the claim by induction on \(s(f)\). If \(s(f) = 0\), then \(\deg f\) is odd and \(f\) has a real root.

\(^{49}\) Also known as the \(d'{Alembert-Gauss\) theorem. Jean le Rond d’Alembert proposes an incomplete proof in 1746. C.F. Gauss gives four correct proofs of this theorem, the first one in 1797. Other proofs are due to Jean Argand (1814), Augustin Louis Cauchy (1820). The “theorem of Liouville” (which is due in fact to Cauchy, 1844) –“any holomorphic bounded function on \(\mathbb{C}\) is constant”– proves the theorem in one line. The present proof belongs to Pierre Samuel and has the advantage (?) of being more “algebraic”. Note that all proofs use some analysis, because fundamental (topological) properties of \(\mathbb{R}\) do not possess purely algebraic descriptions. The essential role is played in fact by the order properties of \(\mathbb{R}\).
Let \( s \in \mathbb{N}, s > 0 \). Suppose that any polynomial \( g \in \mathbb{R}[X] \), with \( s(g) < s \), has a complex root. Let \( f \in \mathbb{R}[X], \text{deg } f = n = 2^s m, s(f) = s \). Let \( t_1, \ldots, t_n \) be the roots of \( f \) in a splitting field of \( f \) over \( \mathbb{C}, K = \mathbb{C}(t_1, \ldots, t_n) \). Let \( P \) be the family of subsets of \( \{1, 2, \ldots, n\} \) having two elements, fix \( a \in \mathbb{R} \) and let \( \{i, j\} \in P \). Consider the following element in \( K \):

\[
u_{ij}(a) := t_it_j + a(t_i + t_j) = u_{ij}(a).
\]

Let \( g_a := \prod_{\{i, j\} \in P} (X - u_{ij}(a)), \) a polynomial in \( K[X] \).

The polynomial \( g_a \) has degree \( n(n - 1)/2 \) and (a priori) coefficients in \( K \). We claim \( g_a \) has coefficients in \( \mathbb{R} \). Indeed, the coefficients of \( g_a \) are polynomial expressions of \( t_1, \ldots, t_n \) with real coefficients. To show that these polynomial expressions are real, let us study how they change under a permutation of the roots \( t_1, \ldots, t_n \). If \( \sigma \) is a permutation of the set \( \{1, 2, \ldots, n\} \), the polynomial obtained from \( g_a \) by the action of \( \sigma \) on the roots (i.e. \( t_i \mapsto t_{\sigma(i)}, \forall i \)) is

\[
\prod_{\{i, j\} \in P} (X - u_{\sigma(i)\sigma(j)}(a))
\]

Since \( \{\{\sigma(i), \sigma(j)\} \mid \{i, j\} \in P\} = P \), the polynomial above coincides with \( g_a \). This means that \( g_a \) has coefficients symmetric polynomials in \( t_1, \ldots, t_n \) with real coefficients. Prop. 2.8 implies that \( g_a \in \mathbb{R}[X] \).

We have \( \frac{n(n - 1)}{2} = 2^{s-1}m(2^s m - 1) \), so \( s(g_a) = s - 1 \) and the induction hypothesis applies to \( g_a \in \mathbb{R}[X] \). Thus, \( g_a \) has a complex root. On the other hand, the roots of \( g_a \) are \( u_{ij}(a) \), so there exists \( \{i, j\} \in P \), such that \( u_{ij}(a) \in \mathbb{C} \).

The above proves that, for any \( a \in \mathbb{R} \), there exists \( \{i, j\} \in P \), with \( u_{ij}(a) \in \mathbb{C} \). Because \( \mathbb{R} \) is infinite and \( P \) is finite, there exist \( a, b \in \mathbb{R} \), with \( a \neq b \), such that \( u_{ij}(a) \in \mathbb{C} \) and \( u_{ij}(b) \in \mathbb{C} \). It follows immediately that \( s := t_i + t_j \) and \( p := t_it_j \) are complex numbers and so \( t_i, t_j \in \mathbb{C} \), as roots of \( X^2 - sX + p \in \mathbb{C}[X] \).
In the general case of a polynomial $g \in \mathbb{C}[X]$, let $\bar{g}$ be the conjugate of $g$ (the polynomial obtained by the complex conjugation of the coefficients of $g$). An elementary computation shows that $\bar{g} \cdot g \in \mathbb{R}[X]$ and the first part applies. Therefore, there exists $\alpha \in \mathbb{C}$ with $\bar{g}(\alpha) \cdot g(\alpha) = 0$. If $g(\alpha) = 0$, we are done. If $\bar{g}(\alpha) = 0$, then $g$ has the root $\bar{\alpha}$.

Exercises

Throughout the exercises, $K$ is a field and $\Omega$ is an algebraic closure of $K$.

1. Let $R$ be a commutative ring. The following statements are equivalent:
   a) $R$ is a domain.
   b) Any polynomial $p \in R[X]$ of degree $n \geq 1$ has at most $n$ roots in $R$ (the roots being counted with their multiplicity).
   c) Any polynomial of degree 1 in $R[X]$ has at most one root in $R$.

2. (The skew field of quaternions) Let

   \[
   \mathbb{H} = \left\{ \begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix} \in M_2(\mathbb{C}) \mid u, v \in \mathbb{C} \right\}.
   \]

   a) Show that $\mathbb{H}$ is a skew field\(^{50}\) (division ring) with respect to addition and multiplication of matrices.

---

\(^{50}\) The skew field of quaternions was discovered (invented?) by W.R. Hamilton.
IV. Field extensions

b) Let \( 1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ i := \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \ j := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ k := \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \) Show that any element of \( \mathbb{H} \) is written uniquely as \( a1 + bi + cj + dk, \) where \( a, b, c, d \in \mathbb{R}. \)

c) Let \( Q \) be the multiplicative subgroup of \( \mathbb{H}^\ast \) generated by \( i \) and \( j. \) Prove that \( Q \) is not commutative and has 8 elements. Write down the multiplication table of \( Q. \) (\( Q \) is called the quaternion group). Deduce that \( \mathbb{H} \) is a skew field.

d) Prove that \( X^2 + 1 \in \mathbb{H}[X] \) has an infinity of solutions in \( \mathbb{H}. \) Does this contradict 2.3?

e) Prove that \( \mathbb{H} \) is an extension of \( \mathbb{C}. \)

f) Give an example of a countable skew field.

3. Give an example of an extension \( K \subseteq L, \) where \( L \) is algebraically closed and \( L \) is not algebraic over \( K. \)

4. Let \( K \subseteq L \subseteq E \) be a tower of extensions and let \( f \in K[X]. \) If \( E \) is a splitting field of \( f \) over \( K, \) then \( E \) is a splitting field of \( f \) over \( L. \)

5. Let \( K \subseteq L \subseteq E \) be field extensions and let \( f \in K[X]. \) Suppose \( L = K(x_1, \ldots, x_n) \), where \( x_1, \ldots, x_n \) are roots of \( f \) (not necessarily all of them). Show that \( E \) is a splitting field of \( f \) over \( K \) if and only if \( E \) a splitting field of \( f \) over \( L \).

6. Let \( f \in K[X] \) be a polynomial of degree \( n \) and let \( x_1, \ldots, x_n \) be the roots of \( f \) in some extension \( L \) of \( K. \) Show that \( K(x_1, \ldots, x_{n-1}) \) is a splitting field of \( f \) over \( K. \)

7. If the extension \( K \subseteq L \subseteq \Omega \) is such that any nonconstant polynomial in \( K[X] \) splits over \( L, \) then \( L = \Omega. \)

8. Show that \( (Q'_{\mathbb{R}})(i) = Q'_C. \)

9. Let \( \text{char } K = 0 \) and let \( g \in K[X], \) irreducible, \( \deg g \geq 2. \) If \( \alpha, \beta \in \Omega \) are roots of \( g, \) then \( \alpha - \beta \notin K. \) (\( \text{Hint. Suppose } \alpha - \beta \in K \) and let \( \beta = \alpha + b, \) \( b \in K. \) Then \( K(\alpha) = K(\alpha + b). \) Because \( g \) is irreducible, there exists a \( K \)-isomorphism \( \varphi : K(\alpha) \rightarrow K(\beta) = K(\alpha) \) that takes \( \alpha \) to
\[ \beta = \alpha + b. \] Then \( \alpha, \varphi(\alpha), \varphi(\varphi(\alpha)) = \varphi^2(\alpha), \ldots, \varphi^n(\alpha), \ldots \) are roots of \( \varphi \), for any \( n \). Since \( \varphi^n(\alpha) = \alpha + nb \) and \( \text{char } K = 0 \), these are all distinct.

10. Let \( f, g \in K[X] \), nonconstant and let \( \beta \) be a root of \( f \) in \( \Omega \). Then:
   \( f(\varphi(\alpha)) \) is irreducible in \( K[X] \) \iff \( f \) is irreducible in \( K[X] \) and \( g(X) - \beta \) is irreducible in \( K(\beta)[X] \). (Hint. Let \( f = a \prod (X - \beta_i) \), where \( \beta = \beta_1, \ldots, \beta_n \in \Omega \). If \( \alpha \) is a root of \( g(X) - \beta \), then \( f(\varphi(\alpha)) = 0 \) and: \( f(\varphi(X)) \) is irreducible in \( K[X] \) \iff \( [K(\alpha) : K] = \deg f(\varphi(X)) = \deg f \cdot \deg g \). In the tower of extensions \( K \subseteq K(\beta) \subseteq K(\alpha) \), \( [K(\beta) : K] = \deg \text{Irr}(\beta, K) \) and \( [K(\alpha) : K(\beta)] = \deg \text{Irr}(\alpha, K(\beta)) \).

**IV.3 Finite fields**

We now apply the results obtained so far (notably the existence of the algebraic closure of a field) to determine all finite fields. We already know (1.10) that, if \( F \) is a finite field, then:
- \( \text{char } F \) is a prime \( p > 0 \);
- \( F \) is a finite extension of \( \mathbb{F}_p \) (\( = \mathbb{Z}_p \), the field of integers modulo \( p \));
- \( |F| = p^n \), where \( n = [F : \mathbb{F}_p] \).

These facts raise the following questions: Given a prime \( p \) and \( n \in \mathbb{N}^* \), does there exist a finite field with \( p^n \) elements? If it exists, is it unique (up to isomorphism)? How can one construct it?

**3.1 Definition.** Let \( K \) be a commutative ring, \( \text{char } K = p > 0 \), \( p \) a prime. Let:

\[ \varphi : K \rightarrow K, \varphi(x) = x^p, \forall x \in K. \]

Of course, \( \varphi(xy) = \varphi(x) \varphi(y), \forall x, y \in K \). Also:
IV. Field extensions

\[ \varphi(x + y) = (x + y)^p = \sum_{0 \leq i \leq p} \binom{p}{i} x^{p-i} y^i = x^p + y^p, \]

The last equality holds because \( p \) divides the binomial coefficients

\[ \binom{p}{i} = \frac{p(p-1) \ldots (p-i+1)}{i!}, \text{ for any } 0 < i < p. \]

Thus, \( \varphi \) is a ring homomorphism, called the Frobenius endomorphism\(^{51}\) of \( K \). The image of \( \varphi \) is \( K^p := \{ x^p \mid x \in K \} \), a subring in \( K \). If \( \text{char } K = 0 \), the Frobenius endomorphism is taken to be the identity function, \( \varphi = \text{id} : K \to K \). We use this endomorphism especially if \( K \) is a field (but also for rings, for example in \( K[X] \), where \( K \) is a field of characteristic \( p \)). The Frobenius endomorphism is the identity in the case of the field \( \mathbb{F}_p \) (by the little Fermat theorem, \( x^p = x \), for any \( x \in \mathbb{F}_p \); see also the proof below).

3.2 Theorem. Let \( p \) be a prime and let \( \Omega \) be an algebraic closure of the field \( \mathbb{F}_p \).

a) Let \( n \in \mathbb{N}^* \) and let \( q := p^n \). There exists a unique subfield \( F \) of \( \Omega \) having \( q \) elements, namely the splitting field of \( X^q - X \) over \( \mathbb{F}_p \). In particular, there exists a finite field with \( p^n \) elements and any two fields with \( p^n \) elements are isomorphic.

b) If \( F \) is a field with \( p^n \) elements and \( K \) is a subfield of \( F \), then \( |K| = p^m \), where \( m \) divides \( n \). Conversely, for any divisor \( m \) of \( n \) there exists a unique subfield of \( F \) with \( p^m =: r \) elements, namely \( K = \{ x \in F \mid x^r = x \} \).

Proof. a) Suppose \( F \subseteq \Omega \) is a field with \( q \) elements. The group \( (F^*, \cdot) \) has \( q - 1 \) elements. Applying the Lagrange theorem for the order of an element in a finite group, we obtain \( x^{q-1} = 1 \), so \( x^q = x \), \( \forall x \in F^* \). The prime subfield of \( F \) is \( \mathbb{F}_p \). Consequently, any element of \( F \) is a root of the polynomial \( X^q - X \), considered with coefficients in

---

\(^{51}\) Ferdinand Georg Frobenius (1849-1917), German mathematician.
This polynomial can have at most \( q \) distinct roots in \( \Omega \); since \( |F| = q \), the elements of \( F \) are exactly the roots of \( X^q - X \) in \( \Omega \). This means that \( F \) is the splitting field of \( X^q - X \) over \( \mathbb{F}_p \) and \( F \) is the unique subfield with \( q \) elements of \( \Omega \).

Let us prove that there exists a field with \( q \) elements \( F \subseteq \Omega \). The argument above says \( F \) must be the splitting field of \( X^q - X \) over \( \mathbb{F}_p \). So, let \( g = X^q - X \). Let \( F \) be the set of all roots of \( g \) in \( \Omega \), \( F = \{x \in \Omega \mid x^q = x\} \). We claim \( F \) is a subfield of \( \Omega \). Indeed, \( \psi : \Omega \to \Omega, \psi(x) = x^q, \forall x \in \Omega \), is a field homomorphism (\( \psi = \varphi^n \) where \( \varphi \) is the Frobenius); then \( F \) is the subfield of all elements fixed by this homomorphism. On the other hand, the number of roots of \( g \) in \( \Omega \) is exactly \( q \): \( g \) has at most \( q \) roots and has no multiple roots, since its derivative is \( qX^{q-1} - 1 = -1 \) (see 2.6). We deduce that \( F \), the splitting field of \( g \) over \( \mathbb{F}_p \), has \( q \) elements.

b) Let \( K \) be a subfield of \( F \) (we suppose that \( F \subseteq \Omega \)). Let \( s := |K| \) and let \( t := [F : K] \). We have \( p^n = |F| = s^t \); since \( p \) is a prime, this is possible only if \( s \) is of the form \( p^m \) and \( p^{mt} = p^n \), which implies \( m|n \). Conversely, if \( m \) divides \( n \), then \( n = mt \) for some \( t \). Let \( \Omega \) be an algebraic closure of \( F \) and let \( \psi : \Omega \to \Omega, \psi(x) = x^r, \forall x \in \Omega \), where \( r = p^m \). Let \( L := \{x \in \Omega \mid x^r = x\} = \{x \in \Omega \mid \psi(x) = x\} \). \( L \) is a subfield with \( r \) elements of \( \Omega \). We show that \( L \subseteq F \). We have \( (\psi \circ \psi)(x) = \psi(x^r) = x^{r^2} \) and, by induction, \( \psi^t(x) = x^{r^t}, \forall x \in \Omega \). So, \( x \in L \Rightarrow \psi(x) = x \Rightarrow \psi^t(x) = x \Rightarrow x^{r^t} = x \). But \( r^t = p^n \), and \( F = \{x \in \Omega \mid x^{p^n} = x\} \); thus, \( x \in L \Rightarrow x \in F \). Any subfield \( E \) of \( F \) with \( r \) elements is equal to \( L \) because all elements of \( E \) satisfy the equation \( x^r = x \).

The finite field with \( p^n \) elements (unique up to an isomorphism) is denoted by \( GF(p^n) \) or \( \mathbb{F}_{p^n} \). The notation \( GF(p^n) \), an acronym for “Galois Field”, honors Évariste Galois, who determined the structure of the finite fields in 1830.
3.3 Proposition. For any \( n \in \mathbb{N}^* \), there exists at least an irreducible polynomial of degree \( n \) in \( \mathbb{F}_p[X] \); for any such polynomial \( f \), \( \mathbb{F}_p[X]/(f) \) is a field with \( p^n \) elements.

Proof. Let \( F \) be the splitting field of \( g = X^q - X \) over \( \mathbb{F}_p \), where \( q = p^n \). The next lemma says that \( (F^*, \cdot) \) is a cyclic group. Let \( \alpha \) be a generator of \( F^* \) and let \( f = \text{Irr}(\alpha, \mathbb{F}_p) \). We have that \( F = \mathbb{F}_p(\alpha) \) and \( \deg f = [\mathbb{F}_p(\alpha) : \mathbb{F}_p] = [F : \mathbb{F}_p] = n \). So, \( f \) is irreducible of degree \( n \) in \( \mathbb{F}_p[X] \). Also, \( \mathbb{F}_p[X]/(f) \cong \mathbb{F}_p(\alpha) \), a field with \( q \) elements.

3.4 Lemma. If \( R \) is a domain and \( G \) is a finite subgroup of \( U(R) \) (the multiplicative group of units of \( R \)), then \( G \) is cyclic. In particular, any finite subgroup of \( K^* \) (where \( K \) is a field) is cyclic.

Proof. Let \( G \) be a finite group with \( n \) elements, \( G \leq (U(R), \cdot) \). Let \( m \) be the exponent of \( G \), (the GCD of the orders of the elements of \( G \)). Lagrange's theorem implies that \( m \mid n \). On the other hand, any element in \( G \) is a root of \( X^m - 1 \), which has at most \( m \) roots in \( R \). Thus, \( n \leq m \) and \( m = n \). Since \( G \) is Abelian, there exists an element \( x \) having order \( m \) in \( G \) (for an elementary proof of this, see exercise 3.2), so \( G \) is cyclic, generated by \( x \).

Another proof. Use the invariant factors theorem, applied to the finite Abelian group \( G \). Suppose \( G \) is not cyclic. Then \( G \cong C_1 \times \cdots \times C_m \), where \( m \geq 2 \) and \( C_1, \ldots, C_m \) are cyclic, \( |C_1| = d_1, \ldots, |C_m| = d_m \) and \( d_1 \mid d_2 \mid \cdots \mid d_m \). Then \( X^{d_m} - 1 \) has \( n \) roots in \( R \), where \( n = d_1 \cdots d_m > d_m \), contradiction.

We prove now that any finite division ring is commutative (it is a field). The proof of this result needs some facts on roots of unity and cyclotomic fields, which have also numerous other applications in algebra, number theory etc.

3.5 Definition. Let \( K \) be a field and let \( n \) be a positive integer. The element \( \zeta \in K \) is called an \( nth \) root of unity in \( K \) if \( \zeta^n = 1 \). Let \( U_n(K) \)
IV.3 Finite fields

\[ U_n := \{ x \in K \mid x^n = 1 \}. \]

Since \( X^n - 1 \) has at most \( n \) roots, \( U_n \) is a subgroup with at most \( n \) elements of the multiplicative group \( (K^*, \cdot) \). The element \( \zeta \in U_n \) is called a primitive \( n \)th root of unity in \( K \) if the order of \( \zeta \) in \( U_n \) is \( n \). \( \zeta \) generates \( U_n \) if \( \zeta^n = 1 \), but \( \zeta^m \neq 1, \forall m < n \). We denote by \( P_n(K) \) (or simply \( P_n \)) the set of all primitive \( n \)th roots of unity in \( K \):

\[ P_n(K) := \{ x \in U_n(K) \mid \text{ord } x = n \}. \]

If \( \zeta \) is a root of unity in some extension of \( K \), the extension \( K \subseteq \Omega \) is called a cyclotomic extension.

3.6 Remark. \( U_n(K) \) is cyclic (as a finite subgroup of \( K^* \)). Thus, there exists a primitive \( n \)th root of unity in \( K \) if and only if \( U_n(K) \) has \( n \) elements.

3.7 Examples. a) \( U_n(\mathbb{Q}) = \{ -1, 1 \} \) if \( n \) is even and \( U_n(\mathbb{Q}) = \{ 1 \} \) if \( n \) is odd.

b) \( U_n(\mathbb{C}) = \{ e^{2\pi i k/n} \mid k \in \{1, \ldots, n\} \} \),

where \( e^{2\pi i k/n} = \cos(2\pi k/n) + i\sin(2\pi k/n) \). There are \( \phi(n) \) complex primitive \( n \)th roots of unity, namely \( e^{2\pi i k/n} \), \( \gcd(k, n) = 1, 1 \leq k < n \). Here \( \phi(n) \) is the number of all integers \( k, \gcd(k, n) = 1, 1 \leq k < n \). The function \( \phi : \mathbb{N}^* \rightarrow \mathbb{N}^* \) is called the Euler phi function. We have \( \phi(n) = |U(\mathbb{Z}/n\mathbb{Z})| \), the number of invertible elements in the ring \( \mathbb{Z}_n \).

c) \( U_4(\mathbb{F}_5) = \mathbb{F}_5^* = \{ 1, 2, 3, 4 \}; \ P_4(\mathbb{F}_5) = \{ 2, 3 \}; \ U_4(\mathbb{F}_7) = \{ 1, 6 \}; \ P_4(\mathbb{F}_7) = \emptyset. \)

3.8 Proposition. Let \( \Omega \) be an algebraic closure of the field \( K \) and let \( n \in \mathbb{N}^* \).

a) \( U_n(\Omega) \) has \( n \) elements if and only if char \( K \) does not divide \( n \).

b) If \( \text{char } K = p > 0 \text{ and } n = p^r m, \text{ with } (p, m) = 1 \), then \( U_n(K) = U_m(K) \).
IV. Field extensions

Proof. a) \( U_n(\Omega) \) has \( n \) elements if and only if \( f = X^n - 1 \) has no multiple roots \( \iff (f', f) = 1 \iff (\mu X^{n-1}, X^n - 1) = 1 \iff n \cdot 1 \neq 0 \iff \text{char } K \) does not divide \( n \).

b) In general, \( m | n \) implies \( U_m \subseteq U_n \). Let \( p^t = q \). If \( x \in U_n \), then \( x^n = (x^m)^q = 1 \). But \( y \mapsto y^q, \forall y \in K \), is a field endomorphism (necessarily injective) of \( K \) (it is a power of the Frobenius endomorphism), so \( y^q = 1 \) implies \( y = 1 \). Thus, \( x \in U_n \) implies \( x^m = 1 \), so \( x \in U_m \).

According to the above, it is sufficient to study \( U_n(K) \) when \( \text{char } K \) does not divide \( n \). In this case, there exists a primitive \( n \)th root of unity in some extension of \( K \).

3.9 Definition. Let \( n \in \mathbb{N}^* \) and let \( K \) be a field containing a primitive \( n \)th root of unity. The polynomial (in \( K[X] \))
\[
\Phi_n := \prod_{\zeta \in P_n(K)} (X - \zeta)
\]
is called the \( n \)th cyclotomic polynomial over \( K \). The roots of \( \Phi_n \) are distinct and they are exactly the primitive \( n \)th roots of unity in \( K \), so \( \deg \Phi_n = \varphi(n) \).

If \( \text{char } K = 0 \) (one usually takes \( K = \mathbb{C} \) to ensure a primitive \( n \)th root of unity exists), then
\[
\Phi_n := \prod_{k < n, (k, n) = 1} (X - e^{2k\pi i/n})
\]
is called the \( n \)th cyclotomic polynomial (dropping any reference to the field).

3.10 Remark. In the complex plane, the complex \( n \)th roots of unity are the vertices of a regular polygon with \( n \) sides inscribed in the unit circle. This justifies the name cyclotomic, of Greek origin and meaning approximately “circle dividing”.

3.11 Lemma. a) Suppose \( K \subseteq L \) is an extension, \( f, g \in K[X] \) and \( h \in L[X] \) such that \( f = gh \). Then \( h \in K[X] \). In other words, if \( g \) divides \( f \) in \( L[X] \), then \( g \) divides \( f \) in \( K[X] \).
b) If \( f \in \mathbb{Q}[X] \) is monic and divides \((\text{in } \mathbb{C}[X])\) a monic polynomial \( u \in \mathbb{Z}[X] \), then \( f \in \mathbb{Z}[X] \) and \( u = fg \), where \( g \in \mathbb{Z}[X] \) and \( g \) is monic.

Proof. a) The division with remainder theorem in \( K[X] \) guarantees the existence of \( q, r \in K[X] \) such that \( f = gq + r \), with \( \deg r < \deg g \). The division with remainder in \( L[X] \) of \( f \) to \( g \) is \( f = gh + 0 \). The uniqueness of the quotient and of the remainder in \( L[X] \) implies that \( r = 0 \) and \( h = q \in K[X] \).

b) By a), \( u = fg \), with \( g \in \mathbb{Q}[X] \). Let \( a \) (respectively \( b \)) be the LCM of denominators of the coefficients of \( f \) (respectively \( g \)). Then \( af, bg \in \mathbb{Z}[X] \) and \( af\cdot bg = abu \). Taking the contents of the polynomials, \( c(af)\cdot c(bg) = ab \), since \( c(u) = 1 \). So \( af = c(af)\cdot f_1, bg = c(bg)\cdot g_1 \), with \( f_1, g_1 \in \mathbb{Z}[X] \), primitive. Thus \( c(af)\cdot f_1\cdot c(bg)\cdot g_1 = abu \). Simplify by \( ab = c(af)\cdot c(bg) \) to obtain \( f_1\cdot g_1 = u \); since \( u \) is monic, and \( f_1, g_1 \in \mathbb{Z}[X] \), \( f_1, g_1 \) have leading coefficients \( 1 \) or \( -1 \). Since \( f \) (which is monic) is associated in divisibility with \( f_1 \), we obtain \( f = \pm f_1 \in \mathbb{Z}[X] \). Likewise, \( g \in \mathbb{Z}[X] \).

3.12 Proposition. Let \( n \in \mathbb{N}^* \) and let \( K \) be a field containing a primitive \( n \)th root of unity. Then:

a) For any \( d \) dividing \( n \), there exists a primitive \( d \)th root of unity in \( K \).

b) \( X^n - 1 = \prod_{d | n} \Phi_d \) holds.

c) The coefficients of \( \Phi_n \) are in the prime subfield \( P \) of \( K \).

d) If \( \text{char } K = 0 \) (for example \( K = \mathbb{C} \)), then \( \Phi_n \in \mathbb{Z}[X] \).

Proof. a) If \( \zeta \) is a primitive \( n \)th root of unity and \( m = n/d \), then \( \zeta^m \) is a primitive \( d \)th root of unity.

b) The polynomials in the equality are monic and have the same roots, namely the \( n \)th roots of unity in \( K \).

c) We prove by induction by \( m \) \((1 \leq m \leq n)\) the statement: "for any \( d \) with \( 1 \leq d \leq m \), \( d \mid n \) implies \( \Phi_d \in P[X] \)."
If $m = 1$, then $\Phi_1 = X - 1 \in P[X]$. Let $1 < m \leq n$. We suppose the claim true for any $q < m$ and we prove for $m$. If $m \nmid n$, there is nothing to prove. If $m \mid n$, then $X^m - 1 = \Phi_m \cdot g$, where $g = \prod_{d \mid m, d < m} \Phi_d \in P[X]$, since by the induction hypothesis $\Phi_d \in P[X], \forall d < m, d \mid m$. Apply now the lemma 3.11.a) to obtain that $\Phi_m$ is also in $P[X]$.

$d)$ Repeat word for word the proof of $c)$ (replace $P$ with $\mathbb{Z}$), to obtain that $X^m - 1 = \Phi_m \cdot g$, where $g \in \mathbb{Z}[X], \text{monic}$. Thus, $\Phi_m \in \mathbb{Q}[X]$. Use lemma 3.11.b) to deduce $\Phi_m \in \mathbb{Z}[X]$. \hfill \Box

One can use the formula $X^n - 1 = \prod_{d \mid n} \Phi_d$ to recursively compute $\Phi_n$ (for this, all $\Phi_d$ are needed, for $d \mid n, d < n$). The Möbius inversion formula (see e.g. SPINDLER [1994]) gives an explicit expression for $\Phi_n$, $\forall n \in \mathbb{N}^*$, in terms of products of polynomials $X^d - 1$ or their inverses.

3.13 Example. Consider the 7th cyclotomic polynomial $\Phi_7$ over an algebraic closure of $\mathbb{F}_2$. From $\Phi_1 \Phi_7 = X^7 - 1$ we deduce $\Phi_7 = X^6 + X^5 + \ldots + 1$. By trial and error (or using a computer) we obtain

$$\Phi_7 = (X^3 + X^2 + 1)(X^3 + X + 1),$$

where the right hand side is the irreducible factorization of $\Phi_7$ in $\mathbb{F}_2[X]$ (the polynomials have degree 3 and no roots in $\mathbb{F}_2$).

So, $\Phi_7$ is reducible in $\mathbb{F}_2[X]$. If $\alpha$ is a primitive 7th root of unity, then $\text{Irr}(\alpha, \mathbb{F}_2) = X^3 + X^2 + 1$ or $X^3 + X + 1$.

$\mathbb{F}_2(\alpha)$ is the field with 8 elements $\mathbb{F}_8$. Any nonzero element of $\mathbb{F}_8$ is a primitive 7th root of unity. Why?

In characteristic 0 the behavior above is not present: the cyclotomic polynomials over $\mathbb{C}$ are irreducible (in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$):
**3.14 Theorem.** Let $n \in \mathbb{N}^*$. Then $\Phi_n$ is irreducible in $\mathbb{Z}[X]$ and it is the minimal polynomial over $\mathbb{Q}$ of any complex primitive $n$th root of unity.

**Proof.** Let $\zeta$ be a primitive complex $n$th root of unity and let $f = \text{Irr}(\zeta, \mathbb{Q})$. We show that $f = \Phi_n$; more precisely, we show that $f$ and $\Phi_n$ have the same roots.

Since $f$ divides $X^n - 1$ in $\mathbb{C}[X]$ and $f$ is monic, 3.11.b) implies that $f \in \mathbb{Z}[X]$.

Any root of $f$ is a primitive $n$th root of unity. Indeed, if $\beta$ is a root of $f$, then $K(\zeta)$ and $K(\beta)$ are $K$-isomorphic via an isomorphism that takes $\zeta$ to $\beta$. In particular, we have, $\forall m \in \mathbb{N}^*, \zeta^m = 1$ if and only if $\beta^m = 1$, so $\text{ord} \zeta = \text{ord} \beta$.

This shows that any root of $f$ is a root of $\Phi_n$, so $f \mid \Phi_n$. We need now to prove that any primitive $n$th root of unity is a root of $f$. We will show that:

If $\alpha \in P_n$ and $f(\alpha) = 0$, then $f(\alpha^p) = 0$, for any prime $p$ not dividing $n$. (*)

Since $f(\zeta) = 0$, an inductive argument based on (*) shows that $f(\zeta^m) = 0$, for any $m$ relatively prime to $n$. Since any primitive $n$th root of unity is of the form $\zeta^m$ for some $m$ coprime with $n$, we obtain that all the primitive $n$th roots of unity are roots of $f$.

In order to prove (*), let $\alpha \in P_n$ and let $p$ be prime, $p \nmid n$. By lemma 3.11.b), $\Phi_n = fg$, with $g \in \mathbb{Z}[X]$, monic. Since $\alpha^p \in P_n$, $f(\alpha^p)g(\alpha^p) = \Phi_n(\alpha^p) = 0$. Suppose by contradiction that $f(\alpha^p) \neq 0$, so $\alpha^p$ is a root of $g$. Thus, $\alpha$ is a root for $g_1 := g(X^p) \in \mathbb{Z}[X]$. Because $f = \text{Irr}(\alpha, \mathbb{Q}), f \mid g_1$ in $\mathbb{Z}[X]$ (by 3.11), so $g_1 = fh$, with $h \in \mathbb{Z}[X]$, monic.

In what follows, for any $q \in \mathbb{Z}[X]$, let $\pi(q)$ denote the polynomial $q$ reduced modulo $p$ ($\pi(q)$ is the polynomial in $\mathbb{Z}_p[X]$ whose coefficients are the images in $\mathbb{Z}_p$ of the coefficients of $q$; in other words, $\pi : \mathbb{Z}[X] \rightarrow \mathbb{Z}_p[X]$ is the unique extension to $\mathbb{Z}[X]$ of the canonical homomorphism $\mathbb{Z} \rightarrow \mathbb{Z}_p$). Let $g = a_0 + a_1X + \ldots + X^n$. Then:
\[ \pi(g^p) = \pi(g)^p = (\pi(a_0) + \pi(a_1)X + \ldots + X^n)^p = \pi(a_0)^p + \pi(a_1)^pX^p + \ldots + X^{pn} = \pi(a_0) + \pi(a_1)X^p + \ldots + X^{pn} = \pi(g_1). \]

Since \( g_1 = fh \), \( \pi(g_1) = \pi(g)^p = \pi(f)\pi(h) \). Thus \( \pi(f) \mid \pi(g)^p \) in \( \mathbb{Z}_p[X] \), so all irreducible divisors of \( \pi(f) \) divide also \( \pi(g) \). Because \( \deg \pi(f) = \deg f \), there exists an irreducible common factor \( h \) of \( \pi(f) \) and \( \pi(g) \), \( \deg h \geq 1 \). But then \( \pi(\Phi_n) = \pi(f)\pi(g) \) is divisible with \( h^2 \), so it has multiple roots. This is absurd, since \( \Phi_n \mid X^n - 1 \), so \( \pi(\Phi_n) \) divides \( \pi(X^n - 1) \), who has no multiple roots: its derivative is \( nX^{n-1} \neq 0 \) in \( \mathbb{Z}_p[X] \) (\( p \nmid n \) implies \( n \) invertible in \( \mathbb{Z}_p \)) and \( \pi(X^n - 1) \) is coprime to \( nX^{n-1} \).

We will encounter cyclotomic extensions again in the study of Galois Theory. We need now some elementary facts on the conjugacy classes of a finite group, which will be used here to prove that any finite division ring is commutative.

### 3.15 Proposition. Let \( G \) be a group and let \( a \in G \).

\( a) \) The set \( C(G) := \{ x \in G \mid xy = yx, \ \forall y \in G \} \) (called the center of \( G \)) is a normal Abelian subgroup of \( G \).

\( b) \) The set \( C(a) := \{ x \in G \mid xa = ax \} \) (called the centralizer of \( a \) in \( G \)) is a subgroup of \( G \).

\( c) \) The “conjugacy relation” in \( G \), defined by: \( \forall x, y \in G, x \sim y \) if and only if exists \( z \in G \) such that \( y = z^{-1}xz \), is an equivalence relation on \( G \). If \( x \sim y \), we say that \( x \) is conjugate with \( y \) (or \( y \) is a conjugate of \( x \)).

\( d) \) Let \( C_a := \{ x \in G \mid x \sim a \} \) (the conjugacy class of the element \( a \)). Then \( |C_a| = [G : C(a)] \) (the index of the subgroup \( C(a) \) in \( G \)).

\( e) \) (conjugacy classes formula) We have:

\[ |G| = |C(G)| + \sum_{a \in S} [G : C(a)], \]

where the sum runs on a system \( S \) of representatives of the conjugacy classes of the elements not contained in the center of \( G \).

**Proof.** \( a), b), c) \) have standard proofs and are left to the reader.
IV.3 Finite fields

**d)** It is clear that \( C_a = \{ z^{-1}az \mid z \in G \} \). Let \( G/C(a) := \{ C(a)x \mid x \in G \} \) be the set of right cosets of the subgroup \( C(a) \) in \( G \). Define \( \varphi : G/C(a) \to C_a \) by \( \varphi(C(a)x) = x^{-1}ax, \forall x \in G. \) \( \varphi \) is correctly defined \((x^{-1}ax \) is independent on the representative \( x \) of the class \( C(a)x \)). Indeed, for any \( x, y \in G \): \( C(a)x = C(a)y \iff xy^{-1} \in C(a) \iff axy^{-1} = xy^{-1}a \iff x^{-1}ax = y^{-1}ay \). The injectivity of \( \varphi \) also follows, since \( x^{-1}ax = y^{-1}ay \) implies \( C(a)x = C(a)y \). Because \( \varphi \) is obviously surjective, \( |C_a| = |G/C(a)| = [G : C(a)]. \)

**e)** \( G \) is the disjoint union of the conjugacy classes. Clearly, \( a \in C(G) \iff C(a) = \{ a \} \iff |C_a| = 1. \) Let \( S \) be a system of representatives as in the statement. Then \( S \cup C(G) \) is a system of representatives for the conjugacy classes, so \( G = \bigcup \{ C_a \mid a \in S \cup C(G) \} = (\bigcup \{ C_a \mid a \in S \}) \cup C(G). \)

The unions are disjoint, so, taking cardinals and using \( |C_a| = [G : C(a)] \), we obtain the formula.

We can prove now the following celebrated result:

**3.16 Theorem.** (Wedderburn\textsuperscript{52}, 1909) *Any finite division ring is commutative.*

**Proof.** Let \( K \) be a finite division ring. Let \( p = \text{char} \ K > 0. \) Let \( C \) be the center of the ring \( K, C := \{ x \in K \mid xy = yx, \forall y \in K \}. \) One easily verifies that \( C \) is a commutative subring of \( K \), and \( C \) is a field (of characteristic \( p \)). We have thus the tower of extensions of division rings \( \mathbb{Z}_p \subseteq C \subseteq K. \) Let \( m = [C : \mathbb{Z}_p] \) and let \( n = [K : C]. \) Then \( |C| = p^m := q \) and \( |K| = q^n. \) It is sufficient to prove that \( n = 1. \)

Suppose that \( n > 1. \) The center of the group \((K^*, \cdot)\) is \( C^*. \) For any \( a \in K, \) consider \( \{ x \in K \mid xa = ax \} =: Z(a), \) which is easily seen to be a subfield of \( K \) that includes \( C. \) The centralizer of \( a \) in the group \( K^* \) is

\textsuperscript{52} Joseph Henry Maclagen Wedderburn (1882-1948), Scottish mathematician.
$Z(a)^* := Z(a) \setminus \{0\}$. Let $d(a) = [Z(a) : C]$, so $|Z(a)| = q^{d(a)}$. By 1.27, the multiplicativity of degrees holds also for extensions of division rings, so $\mathbb{Z}_p \subseteq C \subseteq Z(a) \subseteq K$ implies $d_a \mid n$. In the group $(K^*, \cdot)$ apply the conjugacy classes formula:

$$|K^*| = |C^*| + \sum_{a \in S} [K^*: Z(a)^*],$$

$S$ being a system of representatives of the conjugacy classes of elements not contained in $C^*$. Note that $a \in S$ implies $d_a \neq n$ (otherwise $Z(a) = K$, so $a \in C^*$, contradicting the choice of $S$). In the formula above, $|K^*| = q^n - 1$, $|C^*| = q - 1$, $[K^*: Z(a)^*] = \frac{q^n - 1}{q^{d(a)} - 1}$, so:

$$q^n - 1 = q - 1 + \sum_{a \in S} \frac{q^n - 1}{q^{d(a)} - 1}. \quad (1)$$

The cyclotomic polynomial (over $\mathbb{C}$) $\Phi_n$ divides $(X^n - 1)/(X^{d(a)} - 1)$ in $\mathbb{Z}[X]$, $\forall a \in S$. Indeed, 3.12.b) implies $X^n - 1 = \Phi_n \cdot \prod_{d \mid n, d \neq n} \Phi_d$; since $d_a \mid n$, $d_a \neq n$ and $X^{d(a)} - 1 = \prod_{d \mid d(a)} \Phi_d$, the claim follows. So, $\Phi_n(q)$ divides $q^n - 1$ and also divides $(q^n - 1)/(q^{d(a)} - 1)$, $\forall a \in S$; from $\ (1)$ we obtain that $\Phi_n(q) \mid (q - 1)$.

On the other hand, $|\Phi_n(q)| = \prod_{\zeta \in P_n} |q - \zeta|$. We have $|q - \zeta| > q - 1$, $\forall n \geq 2$, $\forall \zeta \in P_n$ (to see this, represent the complex $n$th roots of unity in a plane). So $|\Phi_n(q)| > q - 1$, contradiction with $\Phi_n(q) \mid q - 1$. \qed
Exercises

1. Let \( K \subseteq L \) be an extension of degree \( n \) of finite fields. Show that the lattice of all intermediate fields of the extension is isomorphic to the lattice of the divisors of \( n \) (ordered by divisibility).

2. Let \( G \) be a finite group and let \( \operatorname{exp}(G) := \operatorname{LCM}\{\operatorname{ord} a \mid a \in G\} \) (the exponent of \( G \)). Prove that:
   \begin{enumerate}
   \item If \( a, b \in G \), \( ab = ba \) and \( \operatorname{ord} a, \operatorname{ord} b = 1 \), then \( \operatorname{ord} ab = \operatorname{ord} a \operatorname{ord} b \).
   \item For any \( a, b \in G \), \( ab = ba \) implies that there exists \( c \in G \) with \( \operatorname{ord} c = [\operatorname{ord} a, \operatorname{ord} b] \).
   \item If \( G \) is Abelian, there exists an element \( g \) of \( G \) such that \( \operatorname{ord} g = \operatorname{exp}(G) \).
   \end{enumerate}

3. Let \( K \) be a finite field of characteristic \( p \) and let \( \alpha \in K^* \). If \( \alpha \) generates the group \( K^* \), then \( K = \mathbb{F}_p(\alpha) \). Is the converse true?

4. Let \( K \) be a finite field. Show that any function \( \varphi : K \to K \) is polynomial (there exists a polynomial \( g \in K[X] \) such that \( \varphi(x) = g(x), \forall x \in K \)). More generally, show that any function \( \varphi : K^n \to K \) is polynomial (there is a polynomial \( g \in K[X_1, \ldots, X_n] \) such that \( \varphi(x_1, \ldots, x_n) = g(x_1, \ldots, x_n), \forall (x_1, \ldots, x_n) \in K^n \)). What happens if \( K \) is infinite?

5. Let \( K \) be a finite field with \( q \) elements.
   \begin{enumerate}
   \item Show that, \( \forall n \in \mathbb{N}^* \), the product of all irreducible monic polynomials in \( K[X] \) with degree dividing \( n \) is \( X^{q^n} - X \). (Hint. Show they have the same roots.)
   \item Let \( N(n, q) \) be the number of irreducible monic polynomials in \( K[X] \) having degree \( n \). Then:
     \[ \sum_{d|n} d \cdot N(d, q) = q^n \]
   \item Compute \( N(p, q) \) for \( p \) a prime.
IV. Field extensions

6. Let $K$ be a finite field with $q$ elements. Determine the number of equivalence classes in $M_n(K)$ with respect to similarity, for any $n \leq 3$. (Hint. Use that two matrices are similar $\iff$ they have the same elementary divisors. Make a connection to the $N(r, q)$ for $r \leq n$.)


8. Give an example of a finite field $K$ and of an irreducible polynomial $g \in K[X]$, $\deg g \geq 2$, such that, if $\alpha, \beta$ are roots of $g$ in $\Omega$, then $\alpha - \beta \in K$ (cf. exercise 2.9). (Hint. If $K = \mathbb{F}_p$, $g = X^p - X - a$, with $a \in \mathbb{F}_p$, then $g(X + c) = g(X), \forall c \in \mathbb{F}_p$.)

9. Find the number of irreducible polynomials of degree 2 in $\mathbb{F}_5[X]$. If $\alpha$ is a root of $X^2 + 2 \in \mathbb{F}_5[X]$, then any polynomial of degree 2 in $\mathbb{F}_5[X]$ has a root in $\mathbb{F}_5[\alpha]$.

10. Let $p$ be a prime and let $k \in \mathbb{N}$. Then $\Phi_{pk} = \Phi_p(X^{p^{k-1}})$.

11. Let $\Omega$ be an algebraic closure of the field $\mathbb{F}_p$. Then $\Omega$ has uncountable many subfields. (Hint. For any subset $S$ of the set of all prime numbers $P$, consider the subfield $C_S$ of $\Omega$, $C_S$ is the composite of the fields $\{\mathbb{F}_{p^n} | n$ is a product of primes from $S\}$.)

The following exercises are about the solutions in finite fields of polynomial equations in several unknowns. In the particular case $\mathbb{Z}_p$, these yield results on “congruences mod $p$”. For some developments, see BOREVICH, SHAFAREVICH [1985].

12. Let $F$ be a finite field with $q$ elements and let $r \in \mathbb{N}^\ast$. Show that:

a) $U_r(F) = \{x \in F^\ast | x^r = 1\}$ is a cyclic subgroup of $F^\ast$, with $d$ elements, where $d = (q - 1, r)$.

b) Let $S_r = S_{r, q} = \sum \{x | x \in U_r(F)\}$. Then $S_{r, q} = 1$ if $|U_r(F)| = 1$; $S_{r, q} = 0$ if $|U_r(F)| > 1$. (Hint. If $\alpha$ is a generator of $U_r(F)$, then $\alpha S_r = S_r$).

c) Let $m \in \mathbb{N}^\ast$ and $T_m = T_{m, q} = \sum \{x^m | x \in F\}$. Then $T_{m, q} = -1$ if $(q - 1) \nmid m$; $T_{m, q} = 0$ if $(q - 1) \nmid m$.

d) Let $f \in F[X_1, \ldots, X_n]$, of (total) degree $< n(q - 1)$. Then:
\[ \sum \{ f(x_1, \ldots, x_n) \mid (x_1, \ldots, x_n) \in F^n \} = 0. \]

(Hint. It is sufficient to suppose \( f \) is a monomial. Then the sum above is a product of \( n \) sums of type \( T_m \), and at least one \( m \) is \( < q - 1 \).)

13. (Warning) Let \( F \) be a field with \( q \) elements, \( \text{char} \ F = p \) and \( g \in F[X_1, \ldots, X_n] \), \( \deg g < n \). Then the number \( N \) of solutions in \( F^n \) of the equation \( g(x_1, \ldots, x_n) = 0 \) is a multiple of \( p \). (Hint. Let \( f := 1 - g^{q - 1} \). Then \( \deg f < n(q - 1) \) and \( g(x_1, \ldots, x_n) \neq 0 \iff f(x_1, \ldots, x_n) = 0 \); also \( g(x_1, \ldots, x_n) = 0 \iff f(x_1, \ldots, x_n) = 1 \). Calculate \( \sum \{ f(x_1, \ldots, x_n) \mid (x_1, \ldots, x_n) \in F^n \} \) and use the previous exercise.)

14. (Chevalley) Suppose \( F \) is a field with \( q \) elements, \( \text{char} \ F = p \) and \( g \in \mathbb{Z}_p[X_1, \ldots, X_n] \) is a homogeneous polynomial, \( \deg g < n \). Then the number of solutions in \( F^n \) of the equation \( g(x_1, \ldots, x_n) = 0 \) is a nonzero multiple of \( p \). In particular, the equation has solutions other than \((0, \ldots, 0)\).

### IV.4 Transcendental extensions

Recall that the extension \( K \subseteq L \) is called transcendental if it is not algebraic: \( L \) contains a transcendental element over \( K \). For example, \( \mathbb{R} \) is a transcendental extension of \( \mathbb{Q} \). The field \( K(X) \) of rational fractions with coefficients in \( K \) is a transcendental extension of \( K \). More generally, if \( (X_i)_{i \in I} \) is a family of indeterminates and \( K(X; I) \) is the field of rational fractions in the indeterminates \( (X_i)_{i \in I} \) with coefficients in \( K \), then \( K \subseteq K(X; I) \) is a transcendental extension.

**4.1 Definition.** a) Let \( K \subseteq L \) be a field extension and let \( \{x_1, \ldots, x_n\} \) be a finite subset of \( L \). We call \( \{x_1, \ldots, x_n\} \) an algebraically independent set over \( K \) (or say that \( x_1, \ldots, x_n \) are algebraically independent over
K) if, for any polynomial \( f \in K[X_1, \ldots, X_n] \), \( f(x_1, \ldots, x_n) = 0 \) implies \( f = 0 \). An arbitrary subset \( S \) of \( L \) is called **algebraically independent over** \( K \) if any finite subset of \( S \) is algebraically independent over \( K \). The set \( S \) is called **algebraically dependent over** \( K \) if \( S \) is not algebraically independent over \( K \). This means that there exist \( x_1, \ldots, x_n \in S \), distinct and \( f \in K[X_1, \ldots, X_n] \), with \( f \neq 0 \) and \( f(x_1, \ldots, x_n) = 0 \); we call the equality \( f(x_1, \ldots, x_n) = 0 \) an **algebraic dependence relation over** \( K \) for \( S \).

b) The extension \( K \subseteq L \) is called a **purely transcendental extension** if \( L \) is generated over \( K \) by a set of algebraically independent elements over \( K \): there exists a subset \( S \) of \( L \), algebraically independent over \( K \), such that \( L = K(S) \).

### 4.2 Remark

Let \( (X_s)_{s \in S} \) be a family of indeterminates indexed by \( S \), a subset of \( L \). The subset \( S \) is algebraically independent over \( K \) if and only if the unique homomorphism of \( K \)-algebras \( v : K[X; S] \rightarrow L \) such that \( v(X_s) = s \), \( \forall s \in S \), is injective. The image of this homomorphism is \( K[S] \), the subring generated by \( K \) and \( S \) in \( L \) (the \( K \)-subalgebra of \( L \) generated by \( S \)). Thus, \( S \) is algebraically independent over \( K \) if and only if \( v : K[X; S] \rightarrow K[S] \) \( v(X_s) = s \), \( \forall s \in S \), is an isomorphism. This isomorphism between the domains \( K[X; S] \) and \( K[S] \) induces an isomorphism between the fields of fractions \( K(X; S) \) and \( K(S) \).

In other words, an algebraically independent set over \( K \) behaves the same as a set of indeterminates in the field of fractions of a polynomial ring with coefficients in \( K \). Another way to say this is: the purely transcendental extensions of \( K \) are (up to \( K \)-isomorphism) the extensions \( K \subseteq K(X; I) \), where \( (X_i)_{i \in I} \) is a family of indeterminates.

### 4.3 Examples

a) The singleton \( \{x\} \) is algebraically independent over \( K \) if and only if \( x \) is transcendental over \( K \). Thus, \( \{e\} \) is algebraically independent over \( \mathbb{Q} \). Likewise, \( \{\pi\} \) and \( \{e^3\} \) are algebraically independent sets over \( \mathbb{Q} \). The set \( \{e, e^5\} \) is not algebraically independ-
ent over $\mathbb{Q}$ because $f(e, e^5) = 0$, where $f = Y^5 - X \in \mathbb{Q}[X, Y]$. It is not known if $\{e, \pi\}$ is algebraically independent over $\mathbb{Q}$.

b) The empty set is algebraically independent over any field.

c) Let $K$ be a field and let $n \in \mathbb{N}^*$. The symmetric elementary polynomials $s_1, \ldots, s_n$ are algebraically independent over $K$ in $K(X_1, \ldots, X_n)$. This statement is equivalent to the uniqueness part in the fundamental theorem of symmetric polynomials: “Any symmetric polynomial in $K[X_1, \ldots, X_n]$ is written uniquely as a polynomial with coefficients in $K$ of $s_1, \ldots, s_n$.”

d) In $K \subseteq K(X, Y)$, the elements $X^5$ and $Y$ are algebraically independent over $K$. Prove this. Can you generalize? So, $K[X, Y] \cong K[X^5, Y]$ ($K$-isomorphism), although $K[X^5, Y] \subsetneq K[X,Y].$

4.4 Proposition. Let $K \subseteq L$ be an extension and let $S, T$ be disjoint subsets of $L$, with $S$ algebraically independent over $K$. Then $S \cup T$ is algebraically independent over $K$ if and only if $T$ is algebraically independent over $K$.

Proof. Suppose $S \cup T$ is algebraically independent over $K$. If $T$ is algebraically dependent over $K(S)$, then there exist $t_1, \ldots, t_n \in T$ and a polynomial $f$ in $n$ indeterminates with coefficients in $K[S]$ such that $f(t_1, \ldots, t_n) = 0$. Write down this equality, taking into account the form of the elements in $K[S]$, to obtain an algebraic dependence relation over $K$ for $S \cup T$.

Conversely, let $T$ be algebraically independent over $K(S)$ and suppose $S \cup T$ is algebraically dependent over $K$. Then there exist $s_1, \ldots, s_m \in S$, $t_1, \ldots, t_n \in T$, distinct, and a nonzero polynomial $f$ in $m + n$ indeterminates with coefficients in $K$ such that

$$f(s_1, \ldots, s_m, t_1, \ldots, t_n) = 0.$$ 

Group the terms by the monomials in $t_1, \ldots, t_n$ to obtain a relation of the form $g(t_1, \ldots, t_n) = 0$, with $g$ nonzero polynomial in $n$ indeterminates with coefficients in $K(S)$. 

\[\square\]
4.5 **Definition.** Let $K \subseteq L$ be an extension and let $S$ be a subset of $L$.

a) $S$ is called a **set of algebraic generators of $L$ over $K$** (or say that $S$ **algebraically generates $L$ over $K$**) if $K(S) \subseteq L$ is an algebraic extension. This is equivalent to the fact that any element $\alpha \in L$ satisfies an equation of the form:

$$
\sum_{0 \leq k \leq n} f_k(x_1, \ldots, x_m)\alpha^k = 0,
$$

for some $x_1, \ldots, x_m \in S$ and $f_k \in K[X_1, \ldots, X_m]$, $\forall k \in \{0, \ldots, n\}$, with $f_n(x_1, \ldots, x_m) \neq 0$.

b) $S$ is called a **transcendence basis for the extension $K \subseteq L$** (or a **transcendence basis of $L$ over $K$**) if $S$ is algebraically independent over $K$ and $S$ algebraically generates $L$ over $K$.

4.6 **Examples.** a) The extension $K \subseteq L$ is algebraic if and only if the empty set $\emptyset$ is a transcendence basis of $L$ over $K$.

b) If $(X_i)_{i \in I}$ is a set of indeterminates, then $(X_i)_{i \in I}$ is a transcendence basis of the field of rational fractions $K(X; I)$ over $K$.

c) $\{e\}$ is a transcendence basis of $\mathbb{Q} \subseteq \mathbb{Q}(e, \sqrt{2})$.

4.7 **Proposition.** Let $K \subseteq L$ be an extension and let $S$ be a subset of $L$. The following statements are equivalent:

a) $S$ is a transcendence basis of $L$ over $K$.

b) $S$ is a maximal algebraically independent set over $K$.

c) $S$ is a minimal set of algebraic generators of $L$ over $K$.

**Proof.** $a) \Rightarrow b)$ Let $\alpha \in L \setminus S$. Since $S$ algebraically generates $L$ over $K$, there exist $x_1, \ldots, x_m \in S$ and $f_k \in K[X_1, \ldots, X_m]$ such that $\alpha$ satisfies an equation of the form:

$$
\sum_{0 \leq k \leq n} f_k(x_1, \ldots, x_m)\alpha^k = 0,
$$

so $S \cup \{\alpha\}$ is not algebraically independent over $K$. 
b)⇒a) We must prove that $K(S) \subseteq L$ is algebraic. Suppose $\alpha \in L$ is not algebraic over $K(S)$. Then the previous result shows that $S \cup \{\alpha\}$ is algebraically independent over $K$, contradicting the maximality of $S$.

a)⇒c) We must prove that, for any $\alpha \in S$, $S \setminus \{\alpha\}$ is not a set of algebraic generators of $L$ over $K$, i.e. $L$ is a transcendental extension of $K(S \setminus \{\alpha\})$. Indeed, $\alpha$ is transcendental over $K(S \setminus \{\alpha\})$ (otherwise $S$ would be algebraically dependent over $K$).

c)⇒a) It is sufficient to see that $S$ is algebraically independent over $K$. If it is not so, then there exist $s_1, \ldots, s_n \in S$ and $f \in K[X_1, \ldots, X_n]$, with $f \neq 0$ and $f(s_1, \ldots, s_n) = 0$. Relabeling if necessary, we may suppose that there exists a monomial in $f$ that contains a power of $s_1$. This means that $s_1$ is algebraic over $K(s_2, \ldots, s_n)$, so $S \setminus \{s_1\}$ algebraically generates $L$ over $K$, contradicting the minimality of $S$. □

There is a similarity between the concepts of linear independence (respectively generating set, basis) in vector spaces and algebraic independence (respectively set of algebraic generators, transcendence basis) in field extensions. The same method used at vector spaces (Zorn's Lemma) is used to prove the existence of a transcendence basis of an extension.

4.8 Theorem. Let $K \subseteq L$ be a field extension and let $S, T$ be subsets of $L$ such that $S \subseteq T$, $S$ is algebraically independent over $K$ and $T$ is a set of algebraic generators of $L$ over $K$. Then:

a) There exists a transcendence basis $B$ of $L$ over $K$ such that $S \subseteq B \subseteq T$.

b) The extension $K \subseteq L$ has a transcendence basis.

Proof. a) We look for a transcendence basis as a maximal algebraically independent set $B$ over $K$, with $S \subseteq B \subseteq T$. To this end, define

$$B := \{C \mid S \subseteq C \subseteq T, C \text{ is algebraically independent over } K\}.$$ 

$B$ is nonempty, since $S \in B$. Order $B$ by inclusion. The set $B$ is inductively ordered (straightforward proof) and has thus a maximal
IV. Field extensions

element $B$. We claim that $B$ is a transcendence basis. $B$ is algebraically independent over $K$ since $B \in B$. There is left to show that $L$ is algebraic over $K(B)$. Since $L$ is algebraic over $K(S)$, it is sufficient to prove that any $\alpha$ in $S$ is algebraic over $K(B)$. If $\alpha \in S$ is transcendental over $K(B)$, then $B \cup \{\alpha\}$ is algebraically independent over $K$ (by 4.4), contradicting the maximality of $B$.

b) Apply a) with $S = \emptyset$ and $T = L$. \hfill \Box

Thus, any field extension $K \subseteq L$ can be “decomposed” in a tower of extensions $K \subseteq K(S) \subseteq L$, with $S$ a transcendence basis of $L$ over $K$; $K \subseteq K(S)$ is purely transcendental and $K(S) \subseteq L$ is algebraic.

As the bases in a vector space, any two transcendence bases of an extension have the same cardinal. For the proof, we use the following lemma, which says that, for two transcendence bases, any element in the first base can be replaced with some element in the second base, to obtain a transcendence basis.

4.9 Lemma. Let $K \subseteq L$ be a field extension and let $S$, $T$ be transcendence bases of $K \subseteq L$. Then, for any $s \in S$, there exists $t \in T$ such that $(S \setminus \{s\}) \cup \{t\}$ is a transcendence basis.

Proof. For any $s \in S$ and $t \in T$, let $S_{st} := (S \setminus \{s\}) \cup \{t\}$. Let $s \in S$. Suppose any $t \in T$ is algebraic over $K(S \setminus \{s\})$. Then $L$ is algebraic over $K(S \setminus \{s\})$, because $L$ is algebraic over $K(T)$ and the transitivity of algebraic extension applies. But this contradicts the fact that $S$ is a minimal set of algebraic generators of $L$ over $K$ (see 4.7). Thus, there exists $t \in T$, transcendental over $K(S \setminus \{s\})$. By 4.4, $S_{st}$ is algebraically independent over $K$. Let us show that $S_{st}$ is also a set of algebraic generators of $L$ over $K$. We remark that $s$ is algebraic over $K(S_{st})$. Indeed, if not, $s$ is transcendental over $K(S_{st})$, so $S_{st} \cup \{s\} = S \cup \{t\}$ is algebraically independent over $K$, (use 4.4). But this contradicts the fact that $S$ is a maximal algebraically independent set over $K$. So, any
IV.4 Transcendental extensions

229

element in $S$ is algebraic over $K(S_{st})$. Since $L$ is algebraic over $K(S)$, the transitivity property implies $L$ is algebraic over $K(S_{st})$.

4.10 Theorem. Any two transcendence bases of a field extension $K \subseteq L$ have the same cardinal.

Proof. Suppose first that $K \subseteq L$ has a finite transcendence basis $S = \{s_1, \ldots, s_n\}$. Let $T$ be another transcendence basis of $L$ over $K$. For $s_1 \in S$, the previous lemma yields $t_1 \in T$ such that $S_1 := \{t_1, s_2, \ldots, s_n\}$ is a transcendence basis. Let $1 \leq i < n$ and suppose we have found $t_1, \ldots, t_i \in T$ such that $S_i := \{t_1, t_2, \ldots, t_i, s_{i+1}, \ldots, s_n\}$ is a transcendence basis. Apply again the previous lemma for $S_i$, $T$ and $s_{i+1}$ to obtain $t_{i+1} \in T$ such that $S_{i+1} := \{t_1, t_2, \ldots, t_i, t_{i+1}, s_{i+2}, \ldots, s_n\}$ is a transcendence basis. Thus, there exist $t_1, \ldots, t_n \in T$ such that $S_n := \{t_1, \ldots, t_n\}$ is a transcendence basis. Since $S_n \subseteq T$ and $T$ is also a transcendence basis, we deduce $S_n = T$, so $|T| = |S_n| = n = |S|$.

In the case when any transcendence basis of $L$ over $K$ is infinite, let $S$ and $T$ be two such bases. For any $t \in T$, since $t$ is algebraic over $K(S)$, there exists a finite subset $S_t$ of $S$ such that $t$ is algebraic over $K(S_t)$. Let $S' = \bigcup_{t \in T} S_t$. Since $T$ is algebraic over $K(S')$ and $L$ is algebraic over $K(T)$, $L$ is algebraic over $K(S')$. So $S' = S$, since $S$ is a minimal set of algebraic generators of $L$ over $K$. We have:

$$|S| = |S'| = |\bigcup_{t \in T} S_t| \leq |T \times \mathbb{N}| = |T|.$$  

We used the fact that $T$ is infinite, so $|T \times \mathbb{N}| = |T|$. Thus, $|S| \leq |T|$. By symmetry, $|T| \leq |S|$, so $|S| = |T|$.

This theorem shows that the following definition is correct.

4.11 Definition. Let $K \subseteq L$ be a field extension. The cardinal of a transcendence basis of $L$ over $K$ is called the transcendence degree of the extension $L/K$ and is denoted $\text{trdeg} (L/K)$.
4.12 Example. Let $K$ be a field and let $L := K(X_1, \ldots, X_n)$. Example 4.3 says that the symmetric fundamental polynomials $s_1, \ldots, s_n$ are algebraically independent over $K$, so there exists a transcendence basis of $K \subseteq L$ that includes $\{s_1, \ldots, s_n\}$. On the other hand, $\{X_1, \ldots, X_n\}$ is clearly a transcendence basis of $K \subseteq L$, so $\text{trdeg} (L/K) = n$. This means that $\{s_1, \ldots, s_n\}$ is a transcendence basis of $K(X_1, \ldots, X_n)$ over $K$.

The transcendence degree is “additive” (compare with the property 1.26 of the degree of an extension).

4.13 Proposition. Let $K \subseteq L \subseteq M$ be a tower of field extensions. Then:

$$\text{trdeg} (M/K) = \text{trdeg} (M/L) + \text{trdeg} (L/K).$$

Proof. Let $S$, $T$ be transcendence bases for $K \subseteq L$, respectively $L \subseteq M$. It is clear that $S \cap T = \emptyset$, so it is enough to prove that $S \cup T$ is a transcendence basis for $K \subseteq M$.

Since $T$ is algebraically independent over $L$ and $K(S) \subseteq L$, $T$ is algebraically independent over $K(S)$. By 4.4, $S \cup T$ is algebraically independent over $K$.

In order to prove that $S \cup T$ algebraically generates $M$ over $K$, remark that $M$ is algebraic over $L(T)$. But $L$ is algebraic over $K(S)$, so $L(T)$ is algebraic over $K(S)(T)$. The transitivity of algebraic extensions implies that $M$ is algebraic over $K(S)(T) = K(S \cup T)$.

Exercises

1. Any transcendental extension has an infinity of intermediate fields.
2. Prove that an extension $K \subseteq L$ is finitely generated if and only if its transcendence degree is finite and, if $B$ is a transcendence basis, the
degree \([L : K(B)]\) is finite. (Hint: If \(L = K(S)\), with \(S\) finite, there exists a transcendence basis \(B\) included in \(S\). Then \(L\) is algebraic and finitely generated -by \(S\) over \(K(B)\).

3. Let \(K \subseteq L\) be a field extension and let \(\alpha, \beta \in L\). Then there exists a \(K\)-isomorphism \(K(\alpha) \cong K(\beta)\) that takes \(\alpha\) in \(\beta\) if and only if \(\alpha\) and \(\beta\) are either both transcendental, either both algebraic and have the same minimal polynomial over \(K\).

4. Prove that the following statement is false: “If \(E\) and \(F\) are subfields of the field \(L\) such that \(E \subseteq L\) and \(F \subseteq L\) are algebraic, then \(E \cap F \subseteq L\) is algebraic”. (Hint. \(K(X^2 + X) \cap K(X^2) = K\).)

5. Let \(K \subseteq L\) be a purely transcendental extension. Then \(L\) is not an algebraically closed field. (Hint. Prove that for any field \(K\), \(K(X)\) is not algebraically closed.)

6. Let \(K \subseteq L\) be an extension and let \(\alpha \in L\) be algebraic over \(K\). If \(t \in L\) is transcendental over \(K\), then \(\text{Irr}(\alpha, K) = \text{Irr}(\alpha, K(t))\).

7. Let \(\Omega\) be an algebraically closed field and let \(K\) be a subfield in \(\Omega\) such that \(\text{trdeg}(\Omega/K)\) is infinite. Show that there exists a \(K\)-endomorphism of \(\Omega\) that is not an automorphism.

8. Let \(\Omega\) be an algebraically closed field and let \(K\) be a subfield in \(\Omega\) such that \(\text{trdeg}(\Omega/K)\) is finite. Show that any \(K\)-endomorphism of \(\Omega\) is an automorphism.

9. Let \(\Omega\) be an algebraic closure of \(\mathbb{C}(X)\). Show that there exists a field isomorphism \(\mathbb{C} \cong \Omega\).

10. Let \(K \subseteq C\) be a field extension such that \(C\) is algebraically closed. Let \(\Omega = K'_C\). Prove that, for any \(x_1, \ldots, x_n \in C\), \(K(x_1, \ldots, x_n) \cap \Omega =: L\) is a finite extension of \(K\). (Hint. Let \(B = \{x_1, \ldots, x_r\}\) be a transcendence basis of \(K \subseteq K(x_1, \ldots, x_n)\), with \(r \leq n\). Then \(B\) is a transcendence basis of \(L \subseteq L(x_1, \ldots, x_n) = K(x_1, \ldots, x_n)\).)
V. Galois Theory

The idea behind modern Galois Theory is the following: for a given field extension, one associates to it a group (the Galois group of the extension). Various properties of the extension can then be deduced by investigating its Galois group. The idea of studying a certain structure (in our case, a field extension) by associating to it another structure (in our case, a group) has been very fertile in 20th century mathematics. It can be found in many areas: Algebraic Topology, Class Field Theory (recently generalized in the form of the “Langlands Correspondence”), Algebraic Geometry, Representation Theory, and the list is far from complete. The recent proof of Fermat's Last Theorem\(^1\) uses Galois Theory as a basic tool.

---

\(^1\) Pierre de Fermat (1601-1665), French mathematician. “Fermat’s Last Theorem” claims that if \(n \geq 3\) is an integer, then the equation \(x^n + y^n = z^n\) has no positive integer solutions. All attempts at proving this assertion failed until 1995, although a lot of partial results were proven. In 1995, the English mathematician Andrew Wiles proved the “Shimura-Taniyama conjecture”, a statement which implies FLT (a fact proven by Gerhard Frey and Ken Ribet in 1986). It appeared soon that some parts of the proof were wrong, but A. Wiles and R. Taylor finally gave a proof which is now accepted as correct. A major part of modern number theory and Algebra owes its existence to the efforts of proving FLT.
V.1 Automorphisms

Recall that, given a field extension $K \subseteq L$, a $K$-automorphism of $L$ is a bijective field homomorphism $\sigma : L \to L$ with $\sigma|_K = \text{id}$; in this case, $\sigma^{-1}$ is also a $K$-automorphism of $L$.

1.1 Definition. Let $K \subseteq L$ be a field extension. The Galois group of the extension $K \subseteq L$ is the set of $K$-automorphisms of $L$ (where the group operation is the composition of maps), denoted by $\text{Gal}(L/K)$ or $G(L/K)$.

If $f$ is a polynomial in $K[\mathcal{X}]$ and $F$ is the splitting field of $f$ over $K$, $G(F/K)$ is called the Galois group of the polynomial $f$ over $K$, denoted $G_{f/K}$, or $G_f$, if $K$ is understood.

The transfer of properties from intermediate fields of the extension $K \subseteq L$ to subgroups of the Galois group and vice versa is accomplished by means of two natural maps, called the Galois Connections, which we now define.

We denote by $\text{IF}(L/K)$ the set of intermediate fields of the extension $K \subseteq L$ and by $\text{Subg}(G(L/K))$ the set of subgroups of the group $G(L/K)$.

To any intermediate field $E \in \text{IF}(L/K)$ we associate the subgroup $G(L/E)$, i.e. the set of those $\sigma \in G(L/K)$ for which $\sigma|_E = \text{id}$. Checking that $G(L/E)$ is a subgroup in $G(L/K)$ is trivial. The other way around, to any subgroup $H$ of $G(L/K)$ associate the fixed field of $H$, defined as $\{x \in L \mid \sigma(x) = x, \forall \sigma \in H\}$ and denoted by $L^H$. The proof of the fact that $L^H$ is an intermediate field of $K \subseteq L$ is immediate\(^2\). Thus, we define:

\[^2\text{More generally, for any subset } S \text{ of } G(L/K), \text{ defining } L^S = \{x \in L \mid \sigma(x) = x, \forall \sigma \in S\}, \text{ one can easily verify that } L^S \text{ is a subfield of } L \text{ containing } K \text{ and that } L^S = L^{<S>}, \text{ where } <S> \text{ is the subgroup generated by } S.\]
Φ : IF(L/K) → Subg(G(L/K)), Φ(E) = G(L/E), ∀E ∈ IF(L/K);
Ψ : Subg(G(L/K)) → IF(L/K), Ψ(H) := L^H, ∀H ∈ Subg(G(L/K)).

A natural question arises: Under what conditions the maps Φ and Ψ are bijective and inverse to each other?

The answer is given, for algebraic extensions, by the Fundamental Theorem of Galois Theory. First, let us review some simple but far-reaching properties of K-homomorphisms.

1.2 Proposition. Let K ⊆ L, K ⊆ E be field extensions and ϕ : L → E a K-homomorphism.
a) If L = E and K ⊆ L is finite, then ϕ is an automorphism.
b) For any f ∈ K[X] and any root α of f in L, ϕ(α) is a root of f in E.
c) If S is system of generators of the extension K ⊆ L (that is, L = K(S)), and σ : L → E is a K-homomorphism such that σ|S = ϕ|S, then σ = ϕ.

Proof. a) Viewing L as a K-vector space, ϕ is a K-vector space endomorphism of L. Since ϕ is injective (being a field homomorphism), we have dim Im ϕ = dim L. So Im ϕ = L.
b) Let f = a_0 + a_1X + ... + a_nX^n. Then a_0 + a_1α + ... + a_nα^n = 0. Applying ϕ, it follows that
a_0 + a_1ϕ(α) + ... + a_nϕ(α)^n = ϕ(0) = 0,
since ϕ is a K-homomorphism and the a_i's are in K. Thus, ϕ(α) is a root of f.
c) Any element of L is of the type xy^{-1}, where y ≠ 0, and x, y ∈ K[S], so they are of the form ∑_{(i_1, ..., i_n) ∈ N^n} a_{i_1, ..., i_n} x_1^{i_1} ... x_n^{i_n}, where a_{i_1, ..., i_n} ∈ K and ∑' indicates the fact that the sum has a finite number of terms. For such an element, we have:
\[ \varphi \left( \sum a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \right) = \sum a_{i_1 \ldots i_n} \varphi(x_1)^{i_1} \ldots \varphi(x_n)^{i_n} = \sum a_{i_1 \ldots i_n} \sigma(x_1)^{i_1} \ldots \sigma(x_n)^{i_n} = \sigma \left( \sum a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n} \right), \]

which shows that \( \varphi = \sigma \).

1.3 Corollary. Let \( K \subseteq L \) be a field extension, let \( \sigma \in G(L/K) \) and let \( f \in K[X] \). Then:

a) \( \sigma \) permutes the roots of \( f \) in \( L \).

b) If \( L = K(S) \) (\( S \) generates \( L \) over \( K \)), then \( \sigma \) is determined by its action on \( S \).

c) If \( x_1, \ldots, x_n \) are the distinct roots of \( f \) in an algebraic closure of \( K \) and \( L = K(x_1, \ldots, x_n) \) is the splitting field of \( f \) over \( K \), then \( G(L/K) = G_f \) is isomorphic to a subgroup of the symmetric group \( S_n \).

Proof. c) View \( S_n \) as the group of permutations of the set \( R := \{ x_1, \ldots, x_n \} \). The map \( \varphi : G(L/K) \to S_n, \sigma \mapsto \sigma|_R, \forall \sigma \in G(L/K) \), is a homomorphism (injective, by b)).

These simple facts are fundamental in determining the Galois group of an extension. It is useful to introduce the following concept:

1.4 Definition. Let \( K \subseteq L \) be a field extension and let \( \alpha, \beta \) elements of \( L \), algebraic over \( K \). We say that \( \alpha \) and \( \beta \) are conjugate over \( K \) if they have the same minimal polynomial over \( K \): \( \text{Irr}(\alpha, K) = \text{Irr}(\beta, K) \). In other words, \( \alpha \) and \( \beta \) are roots of the same irreducible polynomial \( f \) with coefficients in \( K \). One also says “\( \beta \) is a conjugate of \( \alpha \)”.

This definition agrees with the terminology “conjugate numbers”, used to designate the complex numbers \( \alpha = a + ib \) and \( \beta = a - ib \) (where \( a, b \in \mathbb{R} \)). Indeed, \( \alpha \) and \( \beta \) have the same minimal polynomial.

---

3 This is in fact the original view that Evariste Galois had on the notion of group associated to a polynomial. We shall exploit this point of view later in order to obtain data on the Galois group of a polynomial.
over $\mathbb{R}$ (find it!). The same remark can be made for the “conjugates” 
$\gamma = a + b \sqrt{d}$ and $\delta = a - b \sqrt{d}$, where $a, b \in \mathbb{Q}$; $\gamma$ and $\delta$ are conjugate 
over $\mathbb{Q}$, in the sense of the above definition.

An algebraic element $\alpha$ over $K$ can have at most $n$ conjugates, 
where $n$ is the degree of $\alpha$ over $K$ ($n = \deg \text{Irr}(\alpha, K)$). The corollary 3.15.a) says that $\sigma(\alpha)$ is a conjugate of $\alpha$, $\forall \sigma \in G(L/K)$.

**1.5 Corollary.** Let $K \subseteq K(\alpha)$ be a simple field extension. Then its 
Galois group $G$ has at most $n$ elements, where $n = [K(\alpha) : K]$ is the de-
gree of $\alpha$.

**Proof.** A $K$-automorphism $\sigma \in G$ is determined by its value in $\alpha$. 
Since $\sigma(\alpha)$ is a conjugate of $\alpha$, there are at most $n$ possibilities to 
choose $\sigma(\alpha)$ from. $\square$

Thus, the elements of the Galois group can be found by inspecting 
their action on a generating set. Moreover, one must remember that 
they transport any element in one of its conjugates. The following 
examples illustrate this technique.

**1.6 Examples.** 
a) Consider the extension $\mathbb{R} \subseteq \mathbb{C}$ and let 
$G = \text{Gal}(\mathbb{C}/\mathbb{R})$. Because $\mathbb{C} = \mathbb{R}[i]$, where $i = \sqrt{-1}$, it is sufficient to 
look for the action of the automorphisms in $G$ on $i$. Because the mini-
mal polynomial of $i$ is $X^2 + 1$, $\sigma(i) \in \{i, -i\}$, $\forall \sigma \in G$. If $\sigma(i) = i$, then 
$\sigma = \text{id}$; if $\sigma(i) = -i$, then $\sigma$ is the “complex conjugation”: 
$\sigma(a + bi) = a + b \sigma(i) = a - bi$, $\forall a, b \in \mathbb{R}$. So, $G(\mathbb{C}/\mathbb{R})$ consists of two 
elements: the identity map and the conjugation map.

b) Let us find $G = G\left(\mathbb{Q}\left(\sqrt[3]{2}\right)/\mathbb{Q}\right)$. For any $\sigma \in G$, $\sigma\left(\sqrt[3]{2}\right)$ is a root of 
$X^3 - 2$; but $\sqrt[3]{2}$ is the only root of $X^3 - 2$ in $\mathbb{Q}\left(\sqrt[3]{2}\right)$, so $\sigma\left(\sqrt[3]{2}\right) = \sqrt[3]{2}$. 
Since $\{ \sqrt[3]{2} \}$ generates the extension, $\sigma = \text{id}$. So, $G = \{\text{id}\}$, but 
$\mathbb{Q} \subsetneq \mathbb{Q}\left(\sqrt[3]{2}\right)$.

c) Let $K = \mathbb{F}_2(X)$ (the rational function field with coefficients in $\mathbb{F}_2$) 
and let $F := \mathbb{F}_2\left(X^2\right)$, the subfield generated by $\mathbb{F}_2$ and $X^2$. Who is
V.1 Automorphisms

Obviously, \( K = F(X) \), where \( X \) is algebraic over \( F \) as a root of
\( h = Y^2 - X^2 \in F[Y] \). This is even the minimal polynomial of \( X \) over \( F \):
otherwise, the minimal polynomial would be of degree 1, which means that \( X \in F \). But \( X \notin F \): indeed, if \( X = f/g \), with \( f, g \in \mathbb{F}_2[X^2] \), then \( Xg = f \). In this equality of polynomials in \( X \), the left hand side has odd degree and the right hand side has even degree, contradiction. Note that \( h \) can be written (in \( K[Y] \)) \( h = (Y - X)^2 \), so \( X \) is a double root of \( h \).

Take now \( \sigma \in G(K/F) \); then \( \sigma(X) \) can only be \( X \) (the unique root of \( h \) in \( K \)). The fact that \( X \) generates \( K \) over \( F \) implies \( \sigma = \text{id} \). So \( G(K/F) = \{ \text{id} \} \), but \( F \subsetneq K \).

In the cases b) and c) above a bijective correspondence between the subgroups of the Galois group and the intermediate fields cannot exist: the Galois group is trivial (thus it has only one subgroup), but the extensions are not trivial (they have at least two intermediate fields).

Let us look closer at the reasons behind this phenomenon. Note that the given extensions are simple (each is generated by a single element \( \alpha \)), so any automorphism \( \sigma \) in the Galois group is perfectly known by its action on \( \alpha \); moreover, \( \sigma(\alpha) \) must be a conjugate of \( \alpha \). But in both cases \( \alpha \) has no conjugates in the extension: in b), the conjugates exist, but they are in a larger extension (for instance, in the splitting field of \( X^3 - 2 \) over \( \mathbb{Q} \)); in c), \( \alpha \) has no conjugates distinct from itself because its minimal polynomial has only one root (which is a double root).

Thus, it appears natural to consider the following two properties that an algebraic extension \( K \subseteq L \) may have:

“Normality”: For any \( \alpha \in L \), \( \text{Irr}(\alpha, K) \) has all its roots in \( L \).

“Separability”: For any \( \alpha \in L \), \( \text{Irr}(\alpha, K) \) has no multiple roots.

It turns out that for finite extensions, these are necessary and sufficient conditions for the bijectivity of the Galois connections. The
following two sections are devoted to the study of the conditions above.

V.2 Normal extensions

2.1 Definition. A field extension $K \subseteq L$ is called a normal extension if it is algebraic and for any $\alpha \in L$, $\text{Irr}(\alpha, K)$ has all its roots in $L$. In this case, one says “$L$ is normal over $K$”.

Remark that the algebraic extension $K \subseteq L$ is normal iff any irreducible polynomial in $K[X]$, which has a root in $L$, has all its roots in $L$. An easy example of normal extension is the extension $K \subseteq \Omega$, where $\Omega$ is an algebraic closure of $K$.

In concrete situations, checking the above definition is rarely practical (for any $\alpha$, one must find $\text{Irr}(\alpha, K)$ and its roots). Before giving other examples, let us give some necessary and sufficient conditions for an algebraic extension to be normal.

2.2 Proposition. Let $K \subseteq L$ be an algebraic extension and let $\Omega$ be an algebraic closure of $L$. The following statements are equivalent:

a) $K \subseteq L$ is normal.

b) For any $K$-homomorphism $\varphi : \Omega \to \Omega$, we have $\varphi(L) \subseteq L$.

c) For any $K$-homomorphism $\varphi : \Omega \to \Omega$, we have $\varphi(L) = L$.

d) There exists a family of polynomials $F \subseteq K[X]$ such that $L$ is the splitting field of $F$ over $K$.

Proof. $a) \Rightarrow b)$ Let $\varphi: \Omega \to \Omega$ be a $K$-homomorphism and take $\alpha \in L$. If $f$ is $\text{Irr}(\alpha, K)$, then $\varphi(\alpha)$ is a root of $f$, so $\alpha \in L$. 
b)⇒a) Fix an $\alpha \in L$ and denote by $f$ its minimal polynomial over $K$. All roots of $f$ lie in $\Omega$; let $\beta \in \Omega$ be another root of $f$. There exists a $K$-isomorphism $\sigma : K(\alpha) \to K(\beta)$ which takes $\alpha$ to $\beta$. Extend $\sigma$ (see theorem IV.2.19) to a $K$-homomorphism $\tau : \Omega \to \Omega$. Because $\tau(L) \subseteq L$ and $\tau(\alpha) = \sigma(\alpha) = \beta$, we obtain $\beta \in L$. So, all roots of $f$ are in $L$.

a)⇒c) By b), we already know that $\phi(L) \subseteq L$. It remains to show that $\phi$ is surjective. Take $\alpha \in L$, let $f = \text{Irr}(\alpha, K)$ and let $S$ be the set of all roots of $f$ in $\Omega$. We saw that $S \subseteq L$ and that $\phi(S) \subseteq S$. But $\phi$ is injective (it is a field homomorphism) and $S$ is finite, so $\phi|_S : S \to S$ is a bijection. Thus there is a $\beta \in S$ with $\phi(\beta) = \alpha$.

d)⇒b) Let $F = \{\text{Irr}(x, K) \mid x \in L\}$. Because $K \subseteq L$ is normal, the roots of any polynomial in $F$ are in $L$, so the splitting field of $F$ over $K$ is included in $L$. The fact that $L$ is included in the splitting field of $F$ over $K$ is obvious.

d)⇒b) Let $S$ be the set of the roots in $\Omega$ of the polynomials in $F$, $S = \{\alpha \in \Omega \mid \exists f \in F$ such that $f(\alpha) = 0\}$. By hypothesis, $L = K(S)$. If $\phi : \Omega \to \Omega$ is a $K$-homomorphism, then $\phi(S) \subseteq S$ (by 1.2.b)), so $\phi(L) \subseteq L$.

2.3 Examples. a) The extension $\mathbb{Q}((\sqrt{2})/\mathbb{Q}$ is not normal, because $\text{Irr}(\sqrt{2}, \mathbb{Q}) = X^3 - 2$ has also complex non real roots.

b) If we want a normal extension of $\mathbb{Q}$ that contains $\sqrt{2}$, we must adjoin the other roots of $X^3 - 2$. We obtain the extension $\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}$, where $\omega \in \mathbb{C}$ is a root of the polynomial $X^2 + X + 1$. Indeed, the roots of $X^3 - 2$ are $\sqrt[3]{2}, \sqrt[3]{2}\omega, \sqrt[3]{2}\omega^2$ and so $\mathbb{Q}(\sqrt[3]{2}, \omega) = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}\omega, \sqrt[3]{2}\omega^2)$ is the splitting field of $X^3 - 2$ over $\mathbb{Q}$.

c) Any extension $K \subseteq L$ of degree 2 is normal. Indeed, pick $\alpha \in L \setminus K$ and let $f = \text{Irr}(\alpha, K)$. Since $f$ has degree 2 (because $K(\alpha) = L$) and $f$ is divisible by $X - \alpha$ in $L[X]$, $f$ decomposes in linear factors in $L[X]$. 
In order to prove that a given finite extension $K \subseteq L$ is normal, the most frequent approach is to prove that $L$ is a splitting field for some polynomial in $K[X]$.

2.4 Remark. If $K \subseteq L$ is normal, $\Omega$ is an algebraic closure of $L$ and $\varphi : \Omega \to \Omega$ is a $K$-homomorphism, then $\varphi \upharpoonright L$ is a $K$-automorphism of $L$ (an element of $G(L/K)$). Since $\Omega$ is a normal extension of $K$, every $K$-homomorphism $\varphi : \Omega \to \Omega$ is an automorphism. So, in the characterizations above, “$K$-homomorphism” can be replaced by “$K$-automorphism”.

2.5 Corollary. Let $K \subseteq L$ be a finite extension. Then the extension $K \subseteq L$ is normal iff there exists $f \in K[X]$ such that $L$ is the splitting field of $f$ over $K$.

Proof. Suppose $K \subseteq L$ is normal; let $S = \{x_1, \ldots, x_n\} \subseteq L$ be such that $L = K(S)$. If we take $f = \text{Irr}(x_1, K) \cdot \ldots \cdot \text{Irr}(x_n, K)$, then $L$ is the splitting field of $f$ over $K$. The converse is already proven. \qed

2.6 Definition. If $K \subseteq L$ is algebraic, $\Omega$ is an algebraic closure of $L$ and $\sigma$ is a $K$-automorphism of $\Omega$, then $\sigma(L)$ is called a conjugate extension of $L$ over $K$ in $\Omega$.

The situation in Examples a), b) above can be generalized: if $K \subseteq L$ is not normal, by adjoining to $L$ the roots of the minimal polynomials of a set of generators of $L$ over $K$, one gets a normal extension.

2.7 Proposition. Let $K$ be a field, $\Omega$ an algebraic closure of $K$ and $L$ an extension of $K$, $L \subseteq \Omega$. Then there exists a unique normal extension $N$ of $K$ such that $L \subseteq N \subseteq \Omega$, which is the smallest with this property. More precisely, for any normal extension $K \subseteq F$ with $L \subseteq F \subseteq \Omega$, we have $N \subseteq F$. Also, $N$ is the composite of the conjugates of $L$ over $K$ in $\Omega$:

$$N = K(\bigcup \{ \sigma(L) \mid \sigma \in \text{Aut}_K(\Omega) \})$$
Moreover, if \( K \subseteq L \) is finite, then \( K \subseteq N \) is finite.

**Proof.** Let \( S \subseteq L \) such that \( L = K(S) \). We consider \( I = \{ \text{Irr}(x, K) \mid x \in S \} \); take \( N \) to be the splitting field of the family \( I \) over \( K \). It is clear that \( N \) includes \( L \) and \( N \) is normal over \( K \). If \( L \) is finite over \( K \), we can take \( S \) finite, so \( I \) is also finite. So, \( N \) is a finite extension of \( K \).

If \( F \) is a normal extension of \( K \) with \( L \subseteq F \subseteq \Omega \), then \( S \subseteq F \), so the splitting field of \( \text{Irr}(x, K) \) over \( K \) is included in \( F \), for any \( x \in S \). Thus, the splitting field of \( I \) over \( K \) (that is, \( N \)) is included in \( F \).

The uniqueness of \( N \) results from the minimal condition proved above: if \( M \) is another extension with the same properties, \( M \subseteq N \) and \( N \subseteq M \), so \( M = N \).

Let \( M \) be the composite of the conjugates of \( L \) over \( K \) in \( \Omega \). If \( \tau \) is a \( K \)-automorphism of \( \Omega \), then
\[
\{ \sigma(L) \mid \sigma \in \text{Aut}_K(\Omega) \} = \{ \tau \sigma(L) \mid \sigma \in \text{Aut}_K(\Omega) \}
\]
So, \( \tau(M) = K(\{ \tau \sigma(L) \mid \sigma \in \text{Aut}_K(\Omega) \}) = K(\{ \sigma(L) \mid \sigma \in \text{Aut}_K(\Omega) \}) = M \).

This shows that \( M \) is a normal extension of \( K \). Obviously, \( L \subseteq M \).

On the other hand, any normal extension \( E \) of \( K \), that includes \( L \), has the property that \( \sigma(E) = E \), \( \forall \sigma \in \text{Aut}_K(\Omega) \), so \( \sigma(L) \subseteq \sigma(E) = E \), which means that \( M \subseteq E \). So, \( M \) satisfies the same conditions as \( N \). From the uniqueness of \( N \) it follows that \( M = N \). \( \square \)

The extension \( N \) constructed above is “the smallest” (in the sense of inclusion) normal extension of \( K \) that includes \( L \) and is called the *normal closure (in \( \Omega \)) of the extension \( L/K \) (or, of \( L \) over \( K \)). Since \( N \) is the splitting field over \( K \) of a family of polynomials, it follows that the normal closure does not depend (up to a \( K \)-isomorphism) on the algebraic closure \( \Omega \) that we choose.

Normal extensions are not transitive: \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \) and \( \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[3]{2}) \) are normal, being of degree 2, but \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \) is not normal (why?). We have however the following:
2.8 Proposition. Let $K \subseteq L$ be a normal extension and $E$ an intermediate field. Then $E \subseteq L$ is normal.

Proof. Let $F$ be a family of polynomials over $K$ such that $L$ is the splitting field of $F$ over $K$. Then $L$ is the splitting field of $F$ (viewed as included in $E[X]$) over $E$. □

2.9 Proposition. Let $K \subseteq L$ be an extension and $E$, $F$ algebraic extensions of $K$, included in $L$. Then:
   a) If $K \subseteq E$ is normal, then $F \subseteq FE$ is normal.
   b) If $K \subseteq E$ and $K \subseteq F$ are normal, then $K \subseteq EF$ and $K \subseteq E \cap F$ are normal.

Proof. a) $E$ is a splitting field over $K$ for some family $I \subseteq K[X]$, so $E = K(S)$, where $S$ is the set of all roots in $\Omega$ of the polynomials in $I$. Then $FE = F(K(S)) = F(S)$, so $FE$ is a splitting field over $F$ for $I$ (viewed as a subset of $F[X]$).

b) Let $\Omega$ be an algebraic closure of $L$ and $\phi : \Omega \to \Omega$ be $K$-automorphism. Then $\phi(E) \subseteq E$, $\phi(F) \subseteq F$ (as normal extensions), so $\phi(EF) \subseteq EF$ and $\phi(E \cap F) \subseteq E \cap F$. So, $K \subseteq EF$ and $K \subseteq E \cap F$ are normal. □

One can see that the results on normal extensions can be proven using either one of the characterizations given at 2.2 (using the automorphisms of the algebraic closure or using splitting fields). We propose to the reader to give alternate proofs for the results on normal extensions.

Exercises

In the exercises, $K$ is a field and $\Omega$ is an algebraic closure of $K$. 
1. Let $x, y \in \Omega$. Prove that: $x$ and $y$ are conjugate over $K$ $\iff$ there exists a $K$-homomorphism $\varphi : \Omega \to \Omega$ such that $\varphi(x) = y$ $\iff$ for any normal extension $K \subseteq L$ such that $L \subseteq \Omega$ and $x, y \in L$, there exists a $K$-homomorphism $\varphi : L \to L$ such that $\varphi(x) = y$.

2. Let $K \subseteq L$ be a normal extension, $K \subseteq E \subseteq L$ an intermediate field and $F$ an algebraically closed field. Then any $K$-homomorphism $\varphi : E \to F$ can be extended to a $K$-homomorphism $\psi : L \to F$.

3. Let $x \in \Omega$ and $p, q \in K[X]$.
   a) If $q(x) \neq 0$ then $q(x') \neq 0$, for any conjugate $x'$ of $x$ over $K$.
   b) Suppose $q(x) \neq 0$. Let $g \in K[X]$ such that $g(p(x)/q(x)) = 0$. Then, for any conjugate $x'$ of $x$ over $K$, $g(p(x')/q(x')) = 0$.

4. Let $L$ be the splitting field over $\mathbb{Q}$ of the polynomial $X^4 - 9$. Find the degree and a basis of the extension $\mathbb{Q} \subseteq L$.

5. Find the normal closures for:
   \[ \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{3}), \mathbb{Q} \subseteq \mathbb{Q}(\sqrt[4]{5}), \mathbb{Q} \subseteq \mathbb{Q}(\sqrt{3}, \sqrt[4]{5}), \mathbb{Q} \subseteq \mathbb{Q}(\sqrt[4]{2}) \]

6. For any $\alpha \in \Omega$, denote by $C_\alpha$ the splitting field over $K$ of the polynomial $\text{Irr}(\alpha, K)$. Prove that $\forall \alpha, \beta \in \Omega$, $K(\alpha) \subseteq K(\beta) \Rightarrow C_\alpha \subseteq C_\beta$ and that $K(\alpha) = K(\beta) \Rightarrow C_\alpha = C_\beta$. Give an example to show that $C_\alpha = C_\beta$ does not necessarily imply $K(\alpha) = K(\beta)$.

7. Let $L$ be an extension of $K$, included in $\Omega$. Show that $F := \{ x \in \Omega \mid L \text{ contains all roots of } \text{Irr}(x, K) \}$ is a subfield of $L$, and $F$ is a normal extension of $K$. If $K \subseteq E \subseteq L$ is such that $K \subseteq E$ is normal, then $E \subseteq F$.

8. Let $K \subseteq L$ be a normal extension and $f \in K[X]$ a irreducible monic polynomial. If $g, h \in L[X]$ are irreducible divisors of $f$ (in $L[X]$), then there exists $\varphi \in \text{Aut}_K(L)$ such that $h = \overline{\varphi}(g)$ (where $\overline{\varphi} : L[X] \to L[X]$ is the unique $K$-algebra homomorphism that extends $\varphi$ and takes $X$ to $x$).

9. Give an example showing that the assumption that $K \subseteq L$ is normal cannot be omitted in the previous statement.
10. Show that any extension of finite fields is normal.

11. Give examples of normal extensions of $\mathbb{Q}$ of degrees: 4, 6, and 8.

12. For any $n \in \mathbb{N}^*$, give an example of a normal extension of degree $n$.

13. Let $L, E$ be two distinct extensions of degree 3 of $K$, included in $\Omega$. Then $[LE : K] \in \{6, 9\}$. We have $[LE : K] = 6 \Leftrightarrow L$ and $E$ are conjugate over $K$.

**V.3 Separability**

In this section, we fix a field $K$. It is convenient to suppose (without any loss of generality) that *all algebraic extensions of the field $K$ that we consider are included in $\Omega$, a fixed algebraic closure of $K$.*

**3.1 Definition.** Let $K \subseteq L$ be a field extension.

a) An element $\alpha$ of $L$ is called **separable** over $K$ if it is algebraic over $L$ and $\text{Irr}(\alpha, K)$ has no multiple roots (in $\Omega$).

b) An irreducible polynomial $f \in K[X]$ is called a **separable polynomial** if $f$ has no multiple roots (in $\Omega$).

c) An arbitrary polynomial in $K[X]$ is called **separable** if all its irreducible factors are separable. Note that $f \in K[X]$ is separable if and only if every root of $f$ (in $\Omega$) is separable over $K$.

d) An element of $L$ is called **inseparable over $K$** if it is algebraic and not separable over $K$.

---

4 This terminology, introduced by B. L. van der Waerden, expresses the idea that the roots of $f$ are "separated" (distinct).
e) An irreducible polynomial \( f \in K[X] \) is called an **inseparable polynomial** if it has repeated roots

f) The extension \( K \subseteq L \) is called a **separable extension** if it is algebraic and every element of \( L \) is separable over \( K \) (one also says “\( L \) is separable over \( K \)”).

g) If \( K \subseteq L \) is algebraic and not separable (there exists at least one element of \( L \), inseparable over \( K \)), \( K \subseteq L \) is called an **inseparable extension** (or we say that “\( L \) is inseparable over \( K \)”).

Note that the irreducible polynomial \( f \in K[X] \) is separable iff the number of distinct roots of \( f \) in \( \Omega \) equals its degree.

**3.2 Example.** a) The polynomial \( X^2 + 1 \in \mathbb{Q}[X] \) is separable over \( \mathbb{Q} \); its roots \( i \) and \( -i \) are thus separable over \( \mathbb{Q} \).

The following simple result will be often used. We remark that the converse is also true (to be proven in 3.14).

**3.3 Proposition.** Let \( K \subseteq L \subseteq M \) be algebraic extensions. If \( K \subseteq M \) is separable, then \( K \subseteq L \) and \( L \subseteq M \) are separable.

**Proof.** Clearly, \( K \subseteq L \) is separable. Let us show that any \( \alpha \) in \( M \) is separable over \( L \). We know that \( \text{Irr}(\alpha, K) \) has no multiple roots. On the other hand, \( \text{Irr}(\alpha, L) \) divides \( \text{Irr}(\alpha, K) \) in \( L[X] \), so \( \text{Irr}(\alpha, L) \) cannot have multiple roots.

The next result characterizes the **irreducible separable polynomials**. The characterization of the polynomials having repeated roots using the formal derivative (IV.2.6) is essential in the proof.

**3.4 Proposition.** Let \( f = a_0 + a_1X + \ldots + a_nX^n \), with \( a_n \neq 0 \), be an irreducible polynomial with coefficients in \( K \).

a) \( f \) has repeated roots (is inseparable) iff \( f' = 0 \) (\( f' \) is the formal derivative of \( f \)).

b) If \( \text{char } K = 0 \), then \( f \) is separable (\( f \) has no repeated roots).
c) If \( \text{char } K = p > 0 \), then \( f \) is inseparable iff \( f \in K[X^p] \). Furthermore, there exists \( e \in \mathbb{N}^* \) and an irreducible separable polynomial \( g \in K[X] \), such that \( f = g\left(X^{p^e}\right)\).

**Proof.** a) Suppose \( f \) is inseparable. Let \( d = \text{GCD}(f, f') \). If \( f' \neq 0 \), \( d \) is a divisor of \( f' \) and \( \deg d \leq \deg f' < \deg f \). Since \( f \) is irreducible, we must have \( d = 1 \), so \( f \) has no repeated roots. Conversely, if \( f' = 0 \), then \( d = f \), so \( \deg d \geq 1 \) and \( f \) has multiple roots.

b) If \( \text{char } K \neq 0 \), \( f' = a_1 + \ldots + na_nX^{n-1} \neq 0 \) because \( na_n \neq 0 \).

c) Let \( \text{char } K = p > 0 \) and \( f' = 0 \). The coefficients of \( f' \) are \( ia_i = (i \cdot 1) \cdot a_i, \ i \in \{1, \ldots, n\} \). Since \( ia_i = 0 \), we have \( i \cdot 1 = 0 \) or \( a_i = 0 \). If \( p \nmid i \), then \( i \cdot 1 \neq 0 \), so \( a_i = 0 \). Thus \( a_i = 0, \ \forall i \in \{1, \ldots, n\} \) with \( p \nmid i \). In other words, \( f \in K[X^p] \).

Let \( e \) be the greatest integer such that \( f \in K[X^{p^e}] \). Let \( g \in K[X] \) with \( f = g\left(X^{p^e}\right) \); then \( g \) is irreducible (otherwise \( f \) would be reducible!) and \( g \notin K[X^p] \) (if \( g \in K[X^p] \), then \( f \in K[X^{p+1}] \), contradicting the maximality of \( e \)). So, \( g \) is separable.

The fields that pose no problems of separability are called **perfect**:

3.5 **Definition.** A field \( K \) is called a **perfect field** if any algebraic extension of \( K \) is separable.

Rephrasing, a field \( K \) is perfect iff any irreducible polynomial in \( K[X] \) is separable. Proposition 3.4.b) says that **any field of characteristic 0 is perfect**. So, the study of separability is meaningful only in characteristic \( p > 0 \).

3.6 **Proposition.** Every algebraic extension of a perfect field is a perfect field.

**Proof.** Suppose \( K \subseteq L \) is algebraic and \( K \) is perfect. If \( \alpha \) is an algebraic element over \( L \), then \( \alpha \) is algebraic over \( K \) (by transitivity) and
Irr(\(\alpha, L\)) divides Irr(\(\alpha, K\)) in \(L[X]\). Irr(\(\alpha, K\)) has only simple roots, so Irr(\(\alpha, L\)) has the same property. Therefore, \(\alpha\) is separable over \(L\).

### 3.7 Proposition

The field \(K\) is perfect iff the Frobenius endomorphism is an automorphism (if \(\text{char } K = p > 0\), this means \(K^p = K\), where \(K^p = \{x^p \mid x \in K\}\)).

**Proof.** If \(\text{char } K = 0\), the Frobenius is the identity map and all is evident. Let \(\text{char } K = p > 0\). Suppose that the Frobenius endomorphism \((\varphi : K \to K, \varphi(x) = x^p, \forall x \in K)\) is an automorphism. If, on the contrary, an irreducible inseparable polynomial \(f \in K[X]\) would exist, then \(f \in K[X^p]\). Let

\[
f = \sum_{i=0}^{n} a_i X^{pi}, \quad a_i \in K.
\]

Since \(\varphi\) is surjective, there exist \(b_i \in K\) such that \(a_i = b_i^p, \forall i \in \{0, ..., n\}\), so

\[
f = \sum_{i=0}^{n} b_i^p X^{pi} = \left(\sum_{i=0}^{n} b_i X^i\right)^p, \quad b_i \in K.
\]

(we used the fact that \(g \mapsto g^p\) is an endomorphism of the ring \(K[X]\), of characteristic \(p\)). But this shows that \(f\) is reducible!

Conversely, let \(K\) be a perfect field. It will be enough to show that, if \(\varphi\) is not surjective, then there exists an inseparable algebraic element over \(K\). Let \(a \in K \setminus K^p\). Consider the polynomial \(f = X^p - a \in K[X]\) and a root \(\alpha\) of \(f\) in \(\Omega\). We have \(\alpha \notin K\) (because otherwise \(a \in K^p\)), so \(\deg \text{Irr}(\alpha, K) > 1\). If \(\beta\) is another root of \(f\), then \(\beta^p = \alpha^p = a\). The injectivity of the Frobenius endomorphism of \(\Omega\) implies that \(\beta = \alpha\). So, \(f\) has only one root, with multiplicity \(p\). The minimal polynomial of \(\alpha\) over \(K\) divides \(f\) and has the root \(\alpha\), with multiplicity equal to its degree, so \(\alpha\) is inseparable over \(K\).

### 3.8 Corollary

The finite fields, the fields of characteristic 0 and the algebraically closed fields are perfect.
3.9 Example. The corollary above shows that, in order to find a field that is not perfect, we must look among infinite fields of characteristic $p > 0$. A “natural” example of such a field is $\mathbb{F}_p(X) =: K$. The choice is correct. Indeed, consider $h = Y^p - X \in K[Y]$. We have that: $h$ is irreducible (by Eisenstein's criterion applied for the prime element $X \in \mathbb{F}_p[X]$); $h$ is not separable, because $h \in K[Y^p]$. Denoting by $\alpha$ a root of $h$, the extension $K \subseteq K(\alpha)$ is inseparable.

A $K$-homomorphism of extensions of $K$ carries a root of a polynomial $f \in K[X]$ in another root of $f$. This leads to the idea of using $K$-homomorphisms as a tool in the study of separability.

3.10 Definition. Let $K \subseteq L$ be an algebraic extension. The cardinal of the set $\text{Hom}_K(L, \Omega)$ (the $K$-homomorphisms defined on $L$ with values in $\Omega$) is denoted by $[L : K]_s$ and is called the separable degree of the extension $K \subseteq L$. In other words, $[L : K]_s$ is the cardinal of the set

$$\{ \sigma : L \rightarrow \Omega \mid \sigma \text{ homomorphism, } \sigma|_K = \iota \},$$

where $\iota : K \rightarrow \Omega$ is the canonic inclusion. For simple extensions, another way to characterize the separable degree is given at $c)$ below.

3.11 Proposition. a) Let $K \subseteq L$ be an algebraic extension, $\Omega'$ another algebraic closure of $L$ and $\varphi : K \rightarrow \Omega'$ a field homomorphism\(^5\). Then $\text{Hom}_K(L, \Omega)$ is in a bijective correspondence with the set of homomorphisms from $L$ to $\Omega'$ that extend $\varphi$,

$$P(L/K, \varphi, \Omega') := \{ \eta : L \rightarrow \Omega' \mid \eta|_K = \varphi, \eta \text{ homomorphism} \}.$$

Consequently, $|P(L/K, \varphi, \Omega')| = |\text{Hom}_K(L, \Omega)| = [L : K]_s$ does not depend on $\varphi$ and $\Omega'$.

b) The separable degree is multiplicative: if $K \subseteq L \subseteq M$ are algebraic extensions, then

$$[M : K]_s = [M : L]_s [L : K]_s.$$

\(^5\) Recall that $\Omega$ is a fixed algebraic closure of $K$ that contains $L$. 
c) If \( L = K(\alpha) \) is a simple algebraic extension of \( K \), then \([K(\alpha) : K]_s\) is equal to the number of conjugates of \( \alpha \) in \( \Omega \).

d) If \( \alpha \) is algebraic over \( K \), then \([K(\alpha) : K]_s \leq [K(\alpha) : K]\).

**Proof.** a) We must find a bijection between \( P(L/K, \varphi, \Omega') \) and \( P(L/K, \iota, \Omega) \), where \( \iota : K \to \Omega \) is the canonic inclusion. There exists an isomorphism \( \psi : \Omega \to \Omega' \) which extends \( \varphi \) (by IV.2.19). To every \( \sigma \in P(L/K, \varphi, \Omega') \) we associate \( \psi^{-1} \circ \sigma \), which belongs to \( P(L/K, \iota, \Omega) \): if \( x \in K \), then \( \psi^{-1}(\sigma(x)) = \psi^{-1}(\varphi(x)) = x = \iota(x) \). The other way round, to every \( \eta \in P(L/K, \iota, \Omega) \) we associate \( \psi \circ \eta \in P(L/K, \varphi, \Omega') \). It is easy to see that these maps are inverse one to each other.

b) For the sake of simplifying the notation, denote \( P(L/K, \varphi, \Omega) \) by \( P(L/K, \varphi) \). Thus, \([M : K]_s = |P(M/K, \iota)|\). For any \( \eta \in P(M/K, \iota) \), \( \eta L \in P(L/K, \iota) \); this shows that \( P(M/K, \iota) \) is the union of the sets \( P(M/L, \sigma) \) when \( \sigma \) runs over \( P(L/K, \iota) \). These sets are mutually disjoint and \(|P(M/L, \sigma)| = [M : L]_s, \forall \sigma \in P(L/K, \iota)\), as shown before. Therefore, we can write:

\[ [M : K]_s = |P(M/K, \iota)| = |P(L/K, \iota)| \cdot [M : L]_s = [L : K]_s \cdot [M : L]_s. \]

c) Let \( f = \text{Irr}(\alpha, K) \) and let \( R \) be the set of all conjugates of \( \alpha \) (the roots of \( f \) in \( \Omega \)). If \( \varphi : K(\alpha) \to \Omega \) is a \( K \)-homomorphism, then \( \varphi(\alpha) \) is a root of \( f \); on the other hand, any \( K \)-homomorphism defined on \( K(\alpha) \) is uniquely determined by its action on \( \alpha \). In other words, the map \( \nu : \text{Hom}_K(K(\alpha), \Omega) \to R \), \( \nu(\varphi) = \varphi(\alpha) \) is injective. This map is also surjective: for any \( \beta \in R \), there exists a \( K \)-isomorphism between \( K(\alpha) \) and \( K(\beta) \).

d) \( \text{Irr}(\alpha, K) \) has at most \( \text{deg} \text{Irr}(\alpha, K) \) roots; but \( \text{deg} \text{Irr}(\alpha, K) = [K(\alpha) : K] \). The previous part yields the desired inequality. \( \square \)

**3.12 Proposition.** If \( L = K(\alpha) \) is a simple algebraic extension of \( K \), then the following conditions are equivalent:

a) \([K(\alpha) : K]_s = [K(\alpha) : K]\).

b) \( \alpha \) is separable over \( K \).
c) $K(\alpha)$ is a separable extension of $K$.

**Proof.** $a) \iff b)$ is clear, if we look at $d)$ above, and $c) \implies b)$ is obvious.

$a) \implies c)$ Let $\beta \in K(\alpha)$. We must show that $\beta$ is separable over $K$, which amounts to $[K(\beta) : K]_s = [K(\beta) : K]$. The multiplicativity of the separable degree shows that

$$[K(\alpha) : K]_s = [K(\alpha) : K(\beta)]_s[K(\beta) : K]_s. \quad (\ast)$$

$\alpha$ is separable over $K$, so it is also over $K(\beta)$, which means that $[K(\alpha) : K(\beta)] = [K(\alpha) : K(\beta)]_s$ and $[K(\alpha) : K] = [K(\alpha) : K]_s$. Rewriting $(\ast)$, we get

$$[K(\alpha) : K] = [K(\alpha) : K(\beta)][K(\beta) : K]_s.$$

But $[K(\alpha) : K] = [K(\alpha) : K(\beta)][K(\beta) : K]$; comparing the last two relations, it follows that $[K(\beta) : K]_s = [K(\beta) : K]$. \qed

The characterization of separable *simple* extensions using the separable degree, given above, holds in the general case of *finite* extensions:

**3.13 Proposition.** Let $K \subseteq L$ be a finite extension. Then:

a) $[L : K]_s \leq [L : K]$.

b) $K \subseteq L$ is separable iff $[L : K]_s = [L : K]$.

**Proof.** $a)$ If $K \subseteq L$ is simple, the inequality it is already proven.

Suppose that the inequality holds for any finite extension that has a set of generators with at most $n - 1$ elements and let us prove it for the extension $K \subseteq K(x_1, \ldots, x_n)$. We have:

$$[K(x_1) : K]_s \leq [K(x_1) : K]$$

and

$$[K(x_1)(x_2, \ldots, x_n) : K(x_1)]_s \leq [K(x_1)(x_2, \ldots, x_n) : K(x_1)].$$

So:

$$[K(x_1)(x_2, \ldots, x_n) : K(x_1)]_s[K(x_1) : K]_s \leq [K(x_1)(x_2, \ldots, x_n) : K(x_1)][K(x_1) : K].$$

Since the (separable) degree is multiplicative, we get the conclusion.
Suppose \( K \subseteq L \) is separable. We prove by induction on \( n \) the next Proposition: “For any \( n \in \mathbb{N}^* \), and any separable extension \( K \subseteq L \) of degree \( n \), \( [L : K]_s = [L : K] \)”. The case \( n = 1 \) is trivial. If \( n > 1 \), let \( \alpha \in L \setminus K \). If \( L = K(\alpha) \), \( \alpha \) is separable and the previous proposition applies. If \( K(\alpha) \neq L \), then the extensions \( K \subseteq K(\alpha) \) and \( K(\alpha) \subseteq L \) are separable, with degrees strictly smaller than \( [L : K] = n \). By induction, we get \( [L : K(\alpha)]_s = [L : K(\alpha)] \) and \( [K(\alpha) : K]_s = [K(\alpha) : K] \). Multiplying these equalities gets us to the conclusion.

Suppose now \( [L : K]_s = [L : K] \) and yet there exists \( \alpha \in L \), inseparable over \( K \). Then \( [K(\alpha) : K]_s < [K(\alpha) : K] \). Since \( [L : K(\alpha)]_s \leq [L : K(\alpha)] \), multiplying these relations yields \( [L : K]_s < [L : K] \), contradiction. \( \square \)

3.14 Theorem (transitivity of separable extensions) Let \( K \subseteq L \subseteq M \) be algebraic extensions. If \( K \subseteq L \) and \( L \subseteq M \) are separable, then \( K \subseteq M \) is separable.

Proof. We shall use the characterization of separability with the separable degree in the finite case. Take \( \alpha \in M \) and let us show that \( \alpha \) is separable over \( K \). Let \( a_0, \ldots, a_n \in L \) be the coefficients of \( \text{Irr}(\alpha, L) \) and let \( L' = K(a_0, \ldots, a_n) \). Evidently, \( \text{Irr}(\alpha, L) = \text{Irr}(\alpha, L') \), so \( \alpha \) is separable over \( L' \). Thus \( [L'(\alpha) : L']_s = [L'(\alpha) : L'] \). The finite extension \( K \subseteq L' \) is separable, because \( K \subseteq L \) is separable; so \( [L' : K]_s = [L' : K] \). The multiplication of the last two equalities yields \( [L'(\alpha) : K]_s = [L'(\alpha) : K] \), which means that \( K \subseteq L'(\alpha) \) is separable. So, \( \alpha \) is separable over \( K \). \( \square \)

Notice the resemblance with the proof of the transitivity of algebraic extensions.

3.15 Corollary. Let \( K \subseteq L \) be any extension and let \( A \) be a set of elements in \( L \), separable over \( K \). Then \( K \subseteq K(A) \) is a separable extension.
Proof. Every element in $K(A)$ is a polynomial expression with coefficients in $K$ in a finite set of elements of $A$. Thus, we can suppose that $A$ is finite. Suppose $A = \{x_1, x_2\}$. Since $x_1$ is separable over $K$, $K \subseteq K(x_1)$ is separable (by 3.12). As $x_2$ is separable over $K$, it is separable over $K(x_1)$, so $K(x_1) \subseteq K(x_1)(x_2)$ is separable. By transitivity, $K \subseteq K(x_1, x_2)$ is separable. The case $A = \{x_1, x_2, \ldots, x_n\}, n > 2$, reduces to the already proven case, by induction. 

A remarkable fact, very important in Galois Theory, is that any finite separable extension is simple. The following lemma is the essential step in proving this.

3.16 Lemma. Let $K$ be an infinite field, $L$ an extension of $K$ and $\alpha, \beta \in L$, algebraic over $K$, where $\beta$ is separable over $K$. Then $K \subseteq K(\alpha, \beta)$ is a simple extension (it has a primitive element).

Proof. Let $K(\alpha, \beta) = E$, $f = \text{Irr}(\alpha, K)$, $m = \deg f$, $g = \text{Irr}(\beta, K)$, $n = \deg g$, $\alpha = \alpha_1, \ldots, \alpha_m$ the roots of $f$ in $\Omega$, $\beta = \beta_1, \ldots, \beta_n$ the roots of $g$ in $\Omega$. Since $\beta$ is separable, $\beta_1, \ldots, \beta_n$ are distinct. We claim that there exists $c \in K^*$ such that:

\[ \forall i \in \{1, \ldots, m\}, \forall j \in \{1, \ldots, n\}, \alpha + c\beta = \alpha_i + c\beta_j \iff i = 1 \text{ and } j = 1. \]

Indeed, the condition on $c$ amounts to $c \notin \{(\alpha_i - \alpha)(\beta_j - \beta)\}^{-1} | 1 \leq i \leq m, 2 \leq j \leq n\};$ since $K$ is infinite, and the set above is finite, there exists $c \in K$ satisfying (i). With $c$ chosen this way, let $\gamma = \alpha + c\beta$. We show that $\gamma$ is a primitive element, i.e. $K(\alpha, \beta) = K(\gamma)$. Because $K(\gamma) \subseteq K(\alpha, \beta)$ is evident, it is enough to show that $\beta \in K(\gamma)$ (this will imply also $\alpha = \gamma - c\beta \in K(\gamma)$).

Let $h(X) = f(\gamma - cX)$, a polynomial with coefficients in $K(\gamma)$. The idea is to show that the GCD of $h$ and $g$ is $X - \beta$; the fact that $d = X - \beta \in K(\gamma)[X]$ implies then $\beta \in K(\gamma)$.

We have $h(\beta) = f(\gamma - c\beta) = f(\alpha) = 0$, which means that $X - \beta$ divides $h$ (and $g$, too) in $\Omega[X]$. The polynomials $h$ and $g$ have no other
common roots in $\Omega$ but $\beta$: if $h(\beta_j) = f(\gamma - c\beta_j) = 0$, then $\gamma - c\beta_j$ is a root of $f$, which means it is among $\alpha_1, \ldots, \alpha_m$. But then condition (i) ensures that $j = 1$, so $\beta_j = \beta_1 = \beta$. Thus, in $\Omega[X]$, $\gcd(h, g) = X - \beta$.

But $h, g \in K(\gamma)[X]$, so $\gcd(h, g) = X - \beta \in K(\gamma)[X]$, that is, $\beta \in K(\gamma)$. 

3.17 Theorem. (Primitive element theorem) Any separable finite extension $K \subseteq L$ is simple (it has a primitive element).

Proof. If $K$ is a finite field, then $L$ is also finite. The multiplicative group $L^*$ is cyclic, and any generator of $L^*$ can be taken as a primitive element.

Suppose now that $K$ is infinite. If there exist $x_1, x_2 \in L$ such that $L = K(x_1, x_2)$, the preceding lemma applies. The general case follows by an easy induction argument on $n \in \mathbb{N}^*$ with the property that exist $x_1, \ldots, x_n \in L$ such that $K(x_1, \ldots, x_n) = L$. 

3.18 Remark. For the extensions of the type $K \subseteq K(\alpha, \beta)$, with $K$ infinite, lemma 3.16 gives also a practical procedure to find a primitive element (or, at least, a class of good candidates for it, if the condition (i) is hard to verify).

The remaining results (until the end of the section) are a somewhat deeper study of (in)separability. The reader interested primarily in the Fundamental Theorem of Galois theory may skip directly to the next section.

Recall that every algebraic extension of a field of characteristic 0 is separable; thus, all that follows is relevant only in characteristic $p > 0$.

3.19 Definition. If $K \subseteq L$ is a field extension, the separable closure of $K$ in $L$ is the set $K^s_L := \{x \in L \mid x$ separable over $K\}$. By 3.15, the separable closure of $K$ in $L$ is a subfield of $L$ and a separable extension of $K$. Clearly, $K^s_L$ is “the largest” separable extension of $K$ in-
cluded in $L$, $(K^s_L)$ includes any separable extension $K \subseteq E$ with $E \subseteq L$.

The separable closure of $K$ in $\Omega$ (an algebraic closure of $K$) is called the *separable closure of $K$*, denoted $K^s$. Notice that $K^s$ is the splitting field over $K$ of the family of all separable polynomials in $K[X]$ (thus $K^s$ is unique up to a $K$-isomorphism).

The following concept is, in a certain sense, the opposite of separability:

**3.20 Definition.** An algebraic element $\alpha$ of the extension $K \subseteq L$ is called *purely inseparable over $K$* if $\text{Irr}(\alpha, K)$ has only one root in $\Omega$ ($\alpha$ itself, which has order of multiplicity $\deg \text{Irr}(\alpha, K)$). In other words, $\text{Irr}(\alpha, K)$ is of the form $(X - \alpha)^n$. Evidently, any purely inseparable element is inseparable. An extension $K \subseteq L$ is called *purely inseparable* if any element in $L$ is purely inseparable over $K$.

Directly from the definitions, it follows that: an element is simultaneously purely inseparable and separable over $K$ iff it belongs to $K$.

The following result gives a clearer picture of purely inseparable elements over a field $K$ with $\text{char} K = p > 0$:

**3.21 Proposition.** Let $\text{char} K = p > 0$ and let $\alpha \in \Omega$. The following statements are equivalent:

a) $\alpha$ is purely inseparable over $K$.

b) The only conjugate of $\alpha$ in $\Omega$ is $\alpha$.

c) There exists $e \in \mathbb{N}$ such that $\alpha^{p^e} \in K$.

d) The minimal polynomial of $\alpha$ over $K$ is of the form $(X - \alpha)^{p^e} = X^{p^e} - \alpha^{p^e}$.

e) The separable degree of the extension $K \subseteq K(\alpha)$ is 1.

If $\alpha$ is purely inseparable over $K$, then $\text{Irr}(\alpha, K) = (X - \alpha)^{p^e}$, where $e$ is minimal with $\alpha^{p^e} \in K$. 
V.3 Separability

Proof. a) ⇔ b) just rephrase the definition of pure inseparability. 

a) ⇒ c) and a) ⇒ d). Let \( f := \text{Irr}(\alpha, K) \). Since \( \alpha \) is inseparable over \( K \), \( f \in K[X^{p^e}] \). Let \( n \in \mathbb{N}^* \) be such that \( f = (X - \alpha)^n \). Then there exist \( e, m \in \mathbb{N} \) with \( n = p^e \cdot m \) and \( \text{GCD}(p, m) = 1 \). We have

\[
 f = \left( (X - \alpha)^{p^e} \right)^m = \left( X^{p^e} - \alpha^{p^e} \right)^m,
\]

so the coefficient of \( \left( X^{p^e} \right)^{m-1} \) is \( (m \cdot 1) \alpha^{p^e} \). Since \( m \cdot 1 \neq 0 \) in \( K \) (because \( p \nmid m \)), we get \( \alpha^{p^e} \in K \). The polynomial \( f \) is irreducible in \( K[X] \), so \( m = 1 \); thus \( f = (X - \alpha)^{p^e} \). If \( \alpha^{p^d} \in K \), for some \( d \in \mathbb{N} \), then

\[
 g := X^{p^d} - \alpha^{p^d} = (X - \alpha)^{p^d} \in K[X] \text{ and } g(\alpha) = 0.
\]

So, \( g \) is divisible by \( f = \text{Irr}(\alpha, K) \). Then \( \deg g \geq \deg f \), that is, \( d \geq e \).

c) ⇒ d), d) ⇒ a) Evident.

b) ⇔ e) It follows from 3.11.c): \([K(\alpha) : K]_s \) is the number of conjugates of \( \alpha \) in \( \Omega \).

3.22 Example. There exist algebraic elements that are inseparable, but not purely inseparable. Take \( K = \mathbb{F}_2(\mathcal{X}) \) and \( g = Y^6 - X \in K[Y] \), irreducible (by Eisenstein: \( X \) is prime in the UFD \( \mathbb{F}_2[\mathcal{X}] \)). A root \( \alpha \) of \( g \) is inseparable over \( K \), because \( g' = 0 \). But \( \alpha \) is not purely inseparable over \( K \), since \( g = Y^6 - \alpha^6 \) and 6 is not a power of 2 = char \( K \).

3.23 Definition. Let \( K \subseteq L \) be a field extension. The set

\[
 K^i_L := \{ x \in L \mid x \text{ is purely inseparable over } K \}
\]

is called the purely inseparable closure of \( K \) in \( L \).

3.24 Proposition. Let \( K \subseteq L \) be an algebraic extension, \( \text{char } K = p > 0 \).

a) \( K \subseteq L \) is purely inseparable iff \([L : K]_i = 1 \). In this case, \( K \subseteq L \) is normal and \( \text{Gal}(L/K) = \{ \text{id} \} \).
b) Let $K \subseteq L \subseteq M$ be a tower of extensions. Then: $K \subseteq M$ is purely inseparable iff $K \subseteq L$ and $L \subseteq M$ are purely inseparable.

c) If $K \subseteq L$ is purely inseparable and finite, then $[L : K]$ is a power of $p$.

d) The purely inseparable closure $K^i_L$ of $K$ in $L$ is a subfield of $L$. If $E$ is a purely inseparable extension of $K$, $E \subseteq L$, then $E \subseteq K^i_L$. Moreover, $K^i_L \cap K = K$.

Proof. a) Let $K \subseteq L$ be purely inseparable. For any $\alpha \in L$, $\text{Irr}(\alpha, K)$ has all its roots (namely $\alpha$ itself) in $L$, so $K \subseteq L$ is normal. If $\sigma : L \rightarrow \Omega$ is a $K$-homomorphism, then $\sigma(\alpha)$ is a root of $\text{Irr}(\alpha, K)$, so $\sigma(\alpha) = \alpha, \forall \alpha \in L$. Thus, the canonic inclusion is the only element in $\text{Hom}_K(L, \Omega)$, so $[L : K] = 1$ and $G(L/K) = \{\text{id}\}$. Conversely, if $[L : K] = 1$ and $\alpha \in L$, then $[K(\alpha) : K] \leq [L : K] = 1$, so $\alpha$ is purely inseparable over $K$.

b) Let $K \subseteq M$ be purely inseparable. Then, clearly, $K \subseteq L$ is purely inseparable. Any $\alpha \in M$ has the property that $\alpha^{p^e}$ belongs to $K$ (and to $L$) for some $e$, that is, $\alpha$ is purely inseparable over $L$. Conversely, let $K \subseteq L$ and $L \subseteq M$ be purely inseparable and let $\alpha \in M$. Then there exists $e \in \mathbb{N}$ with $\beta = \alpha^{p^e} \in L$. Since $\beta$ is purely inseparable over $K$, $\beta^{p^d} \in K$ for some $d$. So $\alpha^{p^{e+d}} \in K$.

c) If $L = K(\alpha)$ is a simple extension, then $[L : K] = \deg \text{Irr}(\alpha, K)$, which is a power of $p$. In the general case, $L = K(\alpha_1, \ldots, \alpha_n)$ for some finite set of purely inseparable elements $\alpha_1, \ldots, \alpha_n$ in $L$. Let $K_i = K(\alpha_1, \ldots, \alpha_i), \forall i \leq n$; we have a tower of simple extensions (which are purely inseparable, by b)) $K = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n = L$. For every $i$, the degree of $K_i \subseteq K_{i+1}$ is a power of $p$, so $[L : K]$ is still a power of $p$, being the product of these degrees.

d) Let $\alpha, \beta$ be purely inseparable over $K$, with $\beta \neq 0$. We must show that $\alpha - \beta, \alpha \beta, \beta^{-1}$ are purely inseparable. Let $e, d \in \mathbb{N}$ with $\alpha^{p^e} \in K, \beta^{p^d} \in K$. Then:
\[(\alpha - \beta)^{p^e} = \left(\alpha^{p^e}\right)^{p^d} - \left(\beta^{p^e}\right)^{p^d} \in K.\]

A similar argument works for \(\alpha\beta\) and \(\beta^{-1}\). Since \(K_L^i\) contains all purely inseparable elements over \(K\) in \(L\), \(K_L^i\) includes any purely inseparable extension \(K \subseteq E\) with \(E \subseteq L\). Finally, if \(\alpha\) is separable and purely inseparable over \(K\), then \(\alpha \in K\), so \(K_L^s \cap K_L^i = K\).

\[3.25\textbf{ Proposition.}\] Let \(K \subseteq L\) be an algebraic extension. Then:

a) \(K_L^s \subseteq L\) is purely inseparable.

b) \([L : K]_s = [K_L^s : K]\).

c) \(K_L^i \subseteq L\) is separable iff \(L = K_L^s K_L^i\).

\textbf{Proof.} a) Let \(\alpha \in L\) and let \(g \in K[X]\) be the irreducible separable polynomial such that \(\text{Irr}(\alpha, K)(X) = g(X^{p^e})\) (see 3.4.c)). Let \(a = \alpha^{p^e}\). It follows that \(g(a) = 0\), so \(g = \text{Irr}(a, K)\); thus, \(a\) is separable over \(K\). So \(\alpha^{p^e} = a \in K_L^s\), that is, \(\alpha\) is purely inseparable over \(K_L^s\).

b) By 3.11.b), \([L : K]_s = [K_L^s : K]_s [K_L^s : K]\). Since \(K_L^s \subseteq L\) is purely inseparable, \([L : K_L^s]_s = 1\), so \([L : K]_s = [K_L^s : K]_s = [K_L^s : K]\) (because \(K \subseteq K_L^s\) is separable).

c) Let \(C\) be the composite \(K_L^s K_L^i\). Suppose \(L\) is separable over \(K_L^i\). Then \(L\) is separable over \(C\); \(L\) is also purely inseparable over \(K_L^s\), so it is purely inseparable over \(C\). This shows that \(L = C\).

Suppose now that \(L = K_L^i K_L^s = K_L' (K_L^s)\). This means that \(L\) is generated over \(K_L^i\) by separable elements over \(K\) (which are also separable over \(K_L^i\)). It follows that \(L\) is separable over \(K_L^i\).\]
The proposition above shows that, for every algebraic extension $K \subseteq L$, the separable closure $K_s^L$ is important in the sense that we can decompose $K \subseteq L$ in a tower of extensions: the separable extension $K \subseteq K_s^L$ followed by the purely inseparable extension $K_s^L \subseteq L$. The situation is not symmetric for the purely inseparable closure $K_i^L$, since $K \subseteq K_i^L$ is purely inseparable, but $K_i^L \subseteq L$ is not always separable (see exercise 13).

3.26 Definition. For an algebraic extension $K \subseteq L$, define the inseparable degree of $L$ over $K$ to be $[L : K]_i := [L : K_s^L]$.

3.27 Remark. If $K \subseteq L$ is finite and $\text{char } K = p > 0$, then $[L : K]_i$ is a power of $p$, as $K_s^L \subseteq L$ is purely inseparable. But $[L : K]_i$ is not necessarily equal to the greatest power of $p$ that divides $[L : K]$ and neither equal to $[K_i^L : K]$.

The results proven until now show that the following properties hold:

a) $K \subseteq L$ is finite $\Rightarrow [L : K] = [L : K_s^L] : [L : K]_i$.

b) $K \subseteq L$ is separable $\iff L = K_s^L \iff [L : K]_i = 1$.

c) $K \subseteq L$ is purely inseparable $\iff L = K_i^L \iff [L : K]_s = 1$.

Exercise 1.14 shows that if $K \subseteq L$ is normal, then $K_i^L \subseteq L$ is separable (so $L = K_s^L K_i^L$).

3.28 Proposition. Let $K$ have characteristic $p > 0$ and let $a \in K$. If $a \not\in K^p$, then the polynomial $f = X^{p^m} - a$ is irreducible in $K[X]$ for all $m \geq 1$. Conversely, if $X^{p^m} - a$ is irreducible in $K[X]$ for some $m \geq 1$, then $a \not\in K^p$.

Proof. Let $q := p^m$ and take $\alpha, \beta$ to be roots of $f$ in $\Omega$. Then $\alpha^q = a = \beta^q$, so $\alpha = \beta$ (by the injectivity of the field homomorphism $x \mapsto x^q$). Consequently, $\alpha$ is the only root of $f$ in $\Omega : f = (X - \alpha)^q$. Let $g \in K[X]$ be a monic irreducible factor of $f$. If $h \in K[X]$ is irreducible monic and $h|f$, then $h = g$. Indeed, $g$ and $h$ have $\alpha$ as a common root.
and hence are not mutually prime; being irreducible and monic, they are equal. So \( f = g^t \), where \( t \in \mathbb{N}^* \); \( t \) is a power of \( p \), because \( t \cdot \text{grad} \, g = p^m \). Let \( b := g(0) \in K \). We have \( b^t = f(0) = -a \). If \( t > 1 \), \( -a \in K^p \). Since \( K^p \) is a subfield, \( a \in K^p \), contradiction. It follows that \( t = 1 \), so \( f = g \), which is irreducible.

Conversely, if \( a = b^p \) for some \( b \in K \), then \( f = \left(X^{p^m-1} - b\right)^p \), contradicting that \( f \) is irreducible.

This proposition gives a method to exhibit purely inseparable elements. In addition, any purely inseparable element can be constructed by this method (cf. 3.21).

**Exercises**

In the exercises, \( K \) denotes a field and \( \Omega \) an algebraic closure of \( K \).

1. Let \( g = X^3 + 3X + 1 \in K[X] \) and \( \alpha \) a root of \( g \) in \( \Omega \). Under what conditions is \( K \subseteq K(\alpha) \) separable?

2. Let \( L/K \) be a finite extension. We want to prove that \( L/K \) is simple iff it has a finite number of intermediate fields. Denote by \( S \) the set of all intermediate fields of \( L/K \).

   a) Suppose \( L = K(\alpha) \), with \( \alpha \in L \). Let \( f = \text{Irr}(\alpha, K) \). Show that, for any \( E \in S \), \( \text{Irr}(\alpha, E) \) is a monic divisor (in \( \Omega[X] \)) of \( f \) and that \( E = K(C) \), where \( C \) is the set of coefficients of \( \text{Irr}(\alpha, E) \). Deduce that \( E \mapsto \text{Irr}(\alpha, E) \) is injective and that \( S \) is finite.

   b) Suppose \( S \) is finite and \( L = K(\alpha, b) \). If \( K \) is infinite, show that \( L \) is simple considering the set of intermediate fields \( K(\alpha + bc), c \in K \). If \( L = K(\alpha_1, \ldots, \alpha_n) \), use an induction on \( n \) to prove that \( L/K \) is simple.

   c) Is the statement still true if \( L/K \) is infinite?
3. Let $K \subseteq L$ be a simple extension of degree $n$. Then $K \subseteq L$ has at most $2^{n-1}$ intermediate fields.

4. Let $\text{char } K = p > 0$ and $f \in K[X]$. Suppose $f$ has a repeated root (in $\Omega$). Prove that, if $p \nmid \deg f$, then $f$ is reducible in $K[X]$.

5. Let $\alpha$ be a separable element over $K$ and $g = \text{Irr}(\alpha, K)$. Prove that, if $\alpha$ is a multiple root (having multiplicity $n$) of some $f \in K[X]$, then $g^n | f$. (Hint: Using formal derivatives, show that any conjugate of $\alpha$ is a root of $f$, of multiplicity $\geq n$.)

6. Suppose $K$ is perfect and $f \in K[X]$ has a repeated root $\alpha$ (in $\Omega$), of multiplicity $> (\deg f)/2$. Show that $\alpha \in K$.

7. Let $K \subseteq L$ be a normal extension. Then $K \subseteq K_L^s$ is normal.

8. Let $n \in \mathbb{N}$ and let $K \subseteq L$ be an algebraic separable extension such that $\forall x \in L$, $[K(x) : K] \leq n$. Then $[L : K] \leq n$.

9. Let $K \subseteq L$ be a normal extension. Then every irreducible polynomial in $K[X]$ decomposes in $L[X]$ in a product of irreducible polynomials that have the same degree.

10. Let $K \subseteq L$ be algebraic and let $x, x' \in L$ be conjugate over $K$. Let $y \in K(x)$ and $g \in K[X]$ such that $y = g(x)$. Then $y = g(x') \iff x'$ is conjugate to $x$ over $K(y)$.

11. Suppose $K \subseteq L$ is an algebraic extension and $S$ is an intermediate field with $K \subseteq S$ separable and $S \subseteq L$ purely inseparable. Then $S = K_L^s$.

12. Suppose $\text{char } K = p > 0$ and $K \subseteq L$ is an algebraic extension. Prove that $K_L^s = K$ is equivalent to: $\forall x \in L$, if $x^p \in K$, then $x \in K$.

13. Suppose $F$ is a field of characteristic 2 and put $K = F(X, Y)$. Let $u$ be a root of the polynomial $T^2 + T + X \in K[T]$, $S = K(u)$ and $v = \sqrt{uY}$ (v is a root of $T^2 - uY \in S[T]$). Let $L = S(v)$. Prove that:
   a) $K \subseteq S$ is separable of degree 2 and $S \subseteq L$ is purely inseparable of degree 2. Deduce that $K_L^s = S$.
   b) If $t \in L$ and $t^2 \in K$, then $t \in K$. (Hint. Find a base in $L/K$ and use the form of the elements in $L$).
V.3 Separability

14. Let $K \subseteq L \subseteq E$ be a tower of algebraic extensions, with $L \subseteq E$ normal and $K \subseteq L$ purely inseparable. Then $K \subseteq E$ is normal.

15. Suppose $K \subseteq L$ is a finite extension of characteristic $p > 0$, such that $L^p \subseteq K$.
   a) Prove that $K \subseteq L$ is purely inseparable.
   b) Suppose $\{x_1, \ldots, x_n\} \subseteq L$ is such that
      
      \[ K \subseteq K(x_1) \subseteq K(x_1, x_2) \subseteq \ldots \subseteq K(x_1, \ldots, x_n) = L. \]
      Show that $[L : K] = p^n$ (a set $\{x_1, \ldots, x_n\}$ with these properties is called a $p$-basis of $K \subseteq L$).
   c) Any two $p$-bases of $K \subseteq L$ have the same cardinal (called the $p$-dimension of the extension $K \subseteq L$).
   d) The extension $K \subseteq L$ has a $p$-basis.
   e) $\{x_1, \ldots, x_n\}$ is a $p$-basis $\iff$ \[ \left\{ x_1^{i_1} \cdots x_n^{i_n} \mid i_1, \ldots, i_n \in \{0, \ldots, p-1\} \right\} \] is a linearly independent set over $K$ $\iff$ \forall i \in \{1, \ldots, n\}, \ x_i \notin K(\{x_1, \ldots, x_n\} \setminus \{x_i\})$.

16. Let $\text{char } K = p > 0$ and consider $L = K(X, Y)$, the field of rational functions in two indeterminates $X$ and $Y$ over $K$. Let $F = K(X^p, Y^p)$. Show that:
   a) $F \subseteq L$ is purely inseparable.
   b) $[L : F] = p^2$.
   c) For any $\alpha \in L$, $\alpha^p \in F$.
   d) $F \subseteq L$ is not simple.
   e) Deduce that $F \subseteq L$ has an infinity of intermediate fields. Prove that $\forall \beta, \gamma \in F$, $F(X + \beta Y) = F(X + \gamma Y)$ iff $\beta = \gamma$.
   f) $\{X, Y\}$ is a $p$-basis of $F \subseteq L$.

Can an extension of degree $p$ have an infinity of intermediate fields?

17. Take $p = 2$ and $K = \mathbb{F}_2$ in the previous problem.
a) Show that \( \{1, X, Y, XY\} \) is a basis of the extension \( F \subseteq L \) and write the general form of an element of \( L \).

b) Let \( F \subset E \subset L \) be an intermediate field. Show that there exists a unique polynomial \( P_E \in F_2[X, Y] \), of the form

\[
P_E = Xu + Yv + XYw,
\]

with \( u, v, w \in F_2[X^2, Y^2] \), \((u, v, w) = 1\), such that \( E = F(P_E) \). Show that \( E \mapsto P_E \) establishes a bijection between the set \{\( E \mid F \subset E \subset L \) is an intermediate field\} and the set of the polynomials \( P_E \) of the form above.

18. Give an example of a normal extension \( K \subset L \) such that \( L \) is a splitting field for a polynomial \( f \in K[X] \), but such that there is no \( g \in K[X] \), irreducible, with \( L \) the splitting field of \( g \) over \( K \).

### V.4 The Fundamental Theorem of Galois Theory

We gathered sufficient data on normal extensions and on separable extensions to deal with the problem of bijectivity of the Galois connections we stated at the beginning of the chapter. First, we generalize Corollary 1.5, concerning the order of the Galois group of a finite extension. An essential step in the proof is the Dedekind lemma\(^6\). Notice the linear algebra methods.

4.1 Proposition. a) (Dedekind's Lemma) Suppose \((G, \cdot)\) is a semigroup, \(K\) is a field, \(n \in \mathbb{N}^*\) and \(\sigma_i : G \to (K^*, \cdot), \forall i \in \{1, \ldots, n\}\) are distinct semigroup homomorphisms. If \(\alpha_1, \ldots, \alpha_n \in K\), such that

---

\(^6\)Julius Wihelm Richard Dedekind (1831-1916), German mathematician, one of the creators of algebraic number theory.
\[ \alpha_1 \sigma_1(x) + \ldots + \alpha_n \sigma_n(x) = 0 \]

for any \( x \in G \), then \( \alpha_1 = \ldots = \alpha_n = 0 \).

b) If \( K \subseteq L \) is a finite extension, then \( |G(L/K)| \leq [L : K] \).

**Proof.** a) The statement can be rephrased as follows: in the \( K \)-vector space \( K^G \) of functions defined on \( G \) with values in \( K \), \( \sigma_1, \ldots, \sigma_n \) are linearly independent.

Suppose the statement is false. Relabelling if necessary, there exists \( m \leq n \) and \( \alpha_1, \ldots, \alpha_m \in K \), all nonzero, such that
\[
\alpha_1 \sigma_1(x) + \ldots + \alpha_m \sigma_m(x) = 0, \quad \forall x \in K. \tag{1}
\]

We may even suppose that \( m \) is the smallest having this property, in the sense that any linear dependence relation between \( \sigma_1, \ldots, \sigma_n \) has at least \( m \) terms. Since \( \sigma_1 \neq \sigma_2 \), there exists \( y \in G \) such that \( \sigma_1(y) \neq \sigma_2(y) \). Replacing \( x \) with \( xy \) in (1), we have:
\[
\alpha_1 \sigma_1(xy) + \ldots + \alpha_m \sigma_m(xy) = \alpha_1 \sigma_1(x) \sigma_1(y) + \ldots + \alpha_m \sigma_m(x) \sigma_m(y) = 0 \tag{2}
\]
\[
\alpha_1 \sigma_1(x) \sigma_1(y) + \ldots + \alpha_m \sigma_m(x) \sigma_1(y) = 0, \tag{3}
\]
(3) is obtained by multiplying (1) with \( \sigma_1(y) \). Subtracting (2) from (3) we get
\[
\alpha_2(\sigma_2(y) - \sigma_1(y)) \sigma_2(x) + \ldots + \alpha_m(\sigma_m(y) - \sigma_1(y)) \sigma_m(x) = 0, \quad \forall x \in K.
\]

In this equality, \( \alpha_2(\sigma_2(y) - \sigma_1(y)) \) is nonzero, so we obtained a linear dependence relation with less than \( m \) terms, contradiction with the minimality of \( m \).

b) First, we notice that \( G(L/K) \) is finite. Indeed, let \( [L : K] = n \) and take \( \{x_1, \ldots, x_n\} \) a \( K \)-basis of \( L \). Then any \( \sigma \in G(L/K) \) is determined by its values in \( x_1, \ldots, x_n \). But \( \sigma(x) \) is a conjugate of \( x \), \( \forall x \in L \), and the number of conjugates of \( x \) is finite. So let \( G(L/K) = \{\sigma_1, \ldots, \sigma_m\} \) and suppose, by contradiction, that \( m > n \). Consider the matrix:
\[
A = \begin{bmatrix}
\sigma_1(x_1) & \sigma_1(x_2) & \ldots & \sigma_1(x_n) \\
\sigma_2(x_1) & \sigma_2(x_2) & \ldots & \sigma_2(x_n) \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_m(x_1) & \sigma_m(x_2) & \ldots & \sigma_m(x_n)
\end{bmatrix} \in M_{m,n}(L).
\]
The rank of \( A \) is at most \( n \), so its rows are linearly dependent. Let \( \alpha_1, \ldots, \alpha_m \in L \), not all zero, such that
\[
\alpha_1 \sigma_1(x_i) + \ldots + \alpha_m \sigma_m(x_i) = 0, \quad \forall i \in \{1, \ldots, n\}. \tag{4}
\]

Since any \( x \in L \) is a \( K \)-linear combination of \( x_1, \ldots, x_n \), and \( \sigma_i \) are \( K \)-homomorphisms, (4) holds for any \( x \in L \).

Viewing \( \sigma_i \) as homomorphisms from \( L^* \) to \( L^* \), Dedekind's lemma applies. We get that \( \alpha_1 = \ldots = \alpha_m = 0 \), contradiction. \( \square \)

4.2 Corollary. Let \( K \subseteq L \) be a finite, normal and separable extension. Then

\[ |G(L/K)| = [L : K]. \]

Proof. It is enough to show that \( |G(L/K)| \geq [L : K] =: n \). The extension is finite and separable, so it has a primitive element \( \alpha \). So \( \deg \text{Irr}(\alpha, K) = n \). The extension is normal, so \( \text{Irr}(\alpha, K) \) has \( n \) roots in \( L \). For any root \( \beta \) of \( \text{Irr}(\alpha, K) \), \( [K(\beta) : K] = n \), thus \( K(\beta) = L \). But \( K(\alpha) \) and \( K(\beta) \) are \( K \)-isomorphic by an isomorphism \( \sigma_\beta \) that transports \( \alpha \) to \( \beta \). Summarizing, for every conjugate \( \beta \) of \( \alpha \), we exhibited a \( K \)-isomorphism \( \sigma_\beta : L \to L \), which is evidently an element of \( G(L/K) \). \( \square \)

The next proposition is an important step in proving the fundamental theorem and shows that the converse of the preceding result also holds.

4.3 Proposition. (E. Artin\(^7\)) Suppose \( L \) is a field, \( H \) is a finite subgroup of the group \( \text{Aut}(L) \) of the automorphisms of \( L \) and \( L^H \) is the fixed field of \( H \). Then the extension \( L^H \subseteq L \) is finite, normal, separable, \( [L : L^H] = |H| \) and \( G(L/L^H) = H \).

\(^7\) Emil Artin (1898-1962), Austrian mathematician (he lived most in Germany and the USA).
Proof. Let \( x \in L \). The set \( C_x = \{ \sigma(x) \mid \sigma \in H \} \) is finite and \( |C_x| \leq |H| \). If \( \sigma \in H \), then \( \sigma(C_x) = C_x \). Consider the polynomial

\[
    fx := \prod_{y \in C_x} (X - y).
\]

For any automorphism \( \sigma \in H \), the isomorphism \( \sigma : L[X] \to L[X] \) that extends \( \sigma \) leaves \( fx \) invariant, so the coefficients of \( fx \) are in \( L^H \). Indeed, if \( g \in L^H[X] \) is such that \( g(x) = 0 \), then \( g(\sigma(x)) = 0 \), \( \forall \sigma \in H \), so any root of \( fx \) is a root of \( g \). Thus, \( fx | g \). Note that \( fx \) has all roots in \( L \), and these are distinct. Summarizing, \( L^H \subseteq L \) is algebraic, normal and separable. Let us show that it is finite. The previous arguments show that, for any \( x \in L \), the simple extension \( L^H \subseteq L^H(x) \) is finite and its degree is at most \( |H| \). If \( L^H \subseteq E \) is a finite extension (where \( E \subseteq L \)), it is simple by the primitive element theorem, so \( [E : L^H] \leq |H| \). Suppose that \( L^H \subseteq L \) is infinite. Then, \( \forall x \in L \), \( L^H(x) \neq L \); take then some \( y \in L \setminus L^H(x) \). This means that the sequence \( \{L^H(x_1, \ldots, x_n) : \} \) is strictly increasing. Pick \( n \in \mathbb{N}^* \) such that \( [L^H(x_1, \ldots, x_n) : L^H] > |H| \). We have reached a contradiction, since the finite extension \( L^H \subseteq L^H(x_1, \ldots, x_n) \) is of degree at most \( |H| \), as we saw. In conclusion, \( L^H \subseteq L \) is finite and its degree is at most \( |H| \). As \( H \subseteq G(L/L^H) \) (and \( |G(L/L^H)| = [L : L^H] \) from the preceding Corollary), \( |G(L/L^H)| = [L : L^H] = |H| \) and \( H = G(L/L^H) \).

4.4 Corollary. Let \( K \subseteq L \) be a finite extension with \( |G(L/K)| \geq [L : K] \). Then \( K \subseteq L \) is normal and separable and \( |G(L/K)| = [L : K] \).

Proof. Let \( G = G(L/K) \). We have \( |G| = [L : K] \) from 4.1.b). Let \( K_0 = L^G \). The proposition above ensures that \( K_0 \subseteq L \) is normal, separable and \( [L : K_0] = |G| = [L : K] \). Because \( K \subseteq K_0 \subseteq L \), we have \( K = K_0 \).
Recall that the Galois connections are defined as follows: for an extension $K \subseteq L$, whose Galois group is $G$, denote by $\text{IF}(L/K)$ the set of its intermediate fields and by $\text{Subg}(G)$ the set of subgroups of $G$. Then we define:

$$
\Phi : \text{IF}(L/K) \to \text{Subg}(G), \quad \Phi(E) = G(L/E), \quad \forall E \in \text{IF}(L/K);
$$

$$
\Psi : \text{Subg}(G) \to \text{IF}(L/K), \quad \Psi(H) = \{ x \in L | \sigma(x) = x, \quad \forall \sigma \in H \} := L^H,
\quad \forall H \in \text{Subg}(G).
$$

The result that follows holds for any extension and collects some general properties of Galois connections. Its proof is a mere application of the definitions:

**4.5 Proposition.** Let $K \subseteq L$ be an extension. Then:

a) For any $E \in \text{IF}(L/K)$, $L^{G(L/E)} \supseteq E$.

b) For any $H \in \text{Subg}(G(L/K))$, $G(L/L^H) \supseteq H$.

c) $\Phi$ is inclusion reversing: $\forall E_1, E_2 \in \text{IF}(L/K), \ E_1 \subseteq E_2 \Rightarrow G(L/E_1) \supseteq G(L/E_2)$.

d) $\Psi$ is inclusion reversing: $\forall H_1, H_2 \in \text{Subg}(G(L/K)), \ H_1 \subseteq H_2 \Rightarrow L^{H_1} \supseteq L^{H_2}$.

**Proof.**
a) $L^{G(L/E)} = \{ x \in L | \sigma(x) = x, \quad \forall \sigma \in G(L/E) \} \supseteq E$.

b), c), d) are proposed as exercises. \qed

**4.6 Theorem.** (The fundamental theorem of Galois Theory) Let $K \subseteq L$ be a finite, normal and separable extension of fields. Then the Galois connections are inclusion-reversing maps that are bijective and inverse to each other. Via these maps, the intermediate fields $E$ with $K \subseteq E$ normal correspond to normal subgroups of $G(L/K)$. More precisely:

a) For any intermediate field $K \subseteq E \subseteq L$, $L^{G(L/E)} = E$ and $[L : E] = |G(L/E)|$.

b) For any subgroup $H \leq G(L/K)$, $G(L/L^H) = H$. 

c) For any intermediate field \( K \subseteq E \subseteq L \) with \( K \subseteq E \) normal, \( G(L/E) \) is a normal subgroup in \( G(L/K) \). Besides, \( G(E/K) \) is canonically isomorphic to the factor group \( G(L/K)/G(L/E) \).

d) If \( H \) is a normal subgroup in \( G(L/K) \), then \( L^H \) is a normal extension of \( K \), and \( G(L^H/K) \) is canonically isomorphic to the factor group \( G(L/K)/H \).

**Proof.**

*a*) Let \( E \) be an intermediate field of \( K \subseteq L \). Then \( E \subseteq L \) is finite, normal and separable, so \([L : E] = |G(L/E)|\) by Cor. 4.2. On the other hand, Prop. 4.3 ensures that \([L : L^{G(L/E)}] = |G(L/E)|\). As \( L^{G(L/E)} \supseteq E \) (by 4.5), we have \( E = L^{G(L/E)} \).

*b*) Proved at 4.3.

c) Consider the “restriction” homomorphism

\[
\text{res} : G(L/K) \to G(E/K), \quad \text{res}(\sigma) = \sigma|_E,
\]

for any \( \sigma \in G(L/K) \). We have \( \sigma|_E \in G(E/K) \) because \( K \subseteq E \) is normal. The homomorphism \( \text{res} \) is surjective, since every \( \tau \in G(E/K) \) extends to a \( K \)-homomorphism \( \tau' : L \to \Omega \) (where \( \Omega \) is an algebraic closure of \( L \)), by IV.2.19. But \( \tau'(L) = L \), since \( K \subseteq L \) is normal, so \( \tau' \in G(L/K) \) and \( \text{res}(\tau) = \tau \). The kernel of \( \text{res} \) is a normal subgroup in \( G(L/K) \),

\[
\ker \text{res} = \{ \sigma \in G(L/K) \mid \sigma|_E = \text{id} \} = G(L/E).
\]

Applying the isomorphism theorem, we get \( G(L/K)/G(L/E) \cong G(E/K) \).

d) Let \( \Omega \) be an algebraic closure of \( L \) and \( \eta \) a \( K \)-automorphism of \( \Omega \). We must prove that \( \eta(L^H) \subseteq L^H \). This amounts to show that, \( \forall x \in L^H \), we have \( \tau \eta(x) = \eta(x) \), \( \forall \tau \in H \). First, observe that \( \eta(L) = L \) (because \( K \subseteq L \) is normal), so we can consider \( \eta \) as belonging to \( G(L/K) \). We have \( \tau \eta(x) = \eta(\eta^{-1} \tau \eta)(x) = \eta \sigma(x) = \eta(x) \), where we denoted \( \eta^{-1} \tau \eta \) by \( \sigma \in H \) and used the fact that \( \sigma(x) = x \), \( \forall x \in L^H \). The isomorphism follows by c), keeping in mind that \( H = G(L/L^H) \).

A normal and separable extension is called a **Galois extension**. The fundamental theorem of Galois Theory says that, for a finite Galois extension, the Galois connections are bijective. A natural question
arises: “Which are all field extensions for which the Galois connections are bijective?” The answer to this problem was given in 1951 by the Romanian mathematician Dan Barbilian\(^8\) (in the paper *Soluția exhaustivă a problemei lui Steinitz* (The exhaustive solution of the Steinitz problem), Acad. R.P.R., Stud. Cerc. Mat. 2 (1951), 195-259). His result states that: Any field extension for which the Galois connections are bijective and inverse to each other is a finite Galois extension. Notice that, if we suppose that the extension is finite, the result is a consequence of 4.3.

If \(K \subseteq L\) is finite Galois, \(\text{IF}(L/K)\) and \(\text{Subg}(G(L/K))\) are anti-isomorphic as ordered sets, since the Galois connections are inclusion-reversing bijections. Moreover, they are anti-isomorphic as lattices: 
\[
\sup(E, F) = EF \text{ in } \text{IF}(L/K) \text{ corresponds to } \inf(\Phi(E), \Phi(F)) = \Phi(E) \cap \Phi(F) \in \text{Subg}(L/K),
\]
that is: \(G(L/EF) = G(L/E) \cap G(L/F)\). A similar statement holds for \(\inf(E, F) = E \cap F\) (see problem 7).

We remark that, if \(K \subseteq L\) is Galois of degree \(n\) and \(x \in L\), the conjugates of \(x\) over \(K\) (the roots of \(\text{Irr}(x, K)\)) are exactly \(\{\sigma x | \sigma \in G(L/K)\} =: \{x_1, \ldots, x_m\}\) (where necessarily \(m \mid n\). Why? We have \(m = n\) iff \(L = K(x)\)). So, \(\text{Irr}(x, K) = (X - x_1)\ldots(X - x_m)\).

**4.7 Example.** The extension \(\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3}) =: L\) is Galois: separability is automatic (the characteristic is 0), and normality follows from the fact that \(L\) is a splitting field of \((X^2 - 2)(X^2 - 3)\) over \(\mathbb{Q}\). The degree is 4 (see example IV.1.28.b), so the Galois group \(G\) has 4 elements, by the fundamental theorem. In order to find this group, we look on the action of the automorphisms in \(G\) on the generators \(\sqrt{2}\) and \(\sqrt{3}\). If \(\sigma \in G\), then \(\sigma(\sqrt{2}) \in \{\sqrt{2}, -\sqrt{2}\}\) (these are the roots of

---

\(^8\) Also known as a poet, under the pen name Ion Barbu (1895-1964).
Irr(\(\sqrt{2}\), \(\mathbb{Q}\)) = \(X^2 - 2\); likewise, \(\sigma(\sqrt{3}) \in \{\sqrt{3}, -\sqrt{3}\}\). Since \(G\) has 4 elements, and the number of all possible choices is also 4, the automorphisms in \(G\) are determined by their action on generators, according to the following table:

<table>
<thead>
<tr>
<th></th>
<th>id</th>
<th>(\sigma)</th>
<th>(\tau)</th>
<th>(\eta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sqrt{2})</td>
<td>(\sqrt{2})</td>
<td>(-\sqrt{2})</td>
<td>(\sqrt{2})</td>
<td>(-\sqrt{2})</td>
</tr>
<tr>
<td>(\sqrt{3})</td>
<td>(\sqrt{3})</td>
<td>(-\sqrt{3})</td>
<td>(-\sqrt{3})</td>
<td>(-\sqrt{3})</td>
</tr>
</tbody>
</table>

For instance, \(\tau(\sqrt{2}) = \sqrt{2}\) and \(\tau(\sqrt{3}) = -\sqrt{3}\). Based on the table above one can compile the multiplication table of \(G\).

We remark that \(G \cong \mathbb{Z}_2 \times \mathbb{Z}_2\) (the 4 Klein group), any element of \(G\) being of order 2. The subgroups of \(G\) are \{id\}, \(<\sigma> = \{\text{id, } \sigma\}\), \(<\tau> = \{\text{id, } \tau\}\), \(<\eta> = \{\text{id, } \eta\}\) and \(G\). So, the extension has 3 proper intermediate fields, corresponding to the proper subgroups \(<\sigma>, <\tau>, <\eta>\). On the other hand, one sees immediately that \(\mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{3}), \mathbb{Q}(\sqrt{6})\) are proper distinct intermediate fields, so these are all intermediate fields. We have \(\sigma(\sqrt{3}) = \sqrt{3}\), so \(L^{<\sigma>} \supseteq \mathbb{Q}(\sqrt{3})\); the equality holds since “the degrees match”: \([\mathbb{Q}(\sqrt{3}): \mathbb{Q}] = [L^{<\sigma>} : \mathbb{Q}]\). But \([L^{<\sigma>} : \mathbb{Q}] = [L : \mathbb{Q}]/[L : L^{<\sigma>}] = 4/2 = 2 = [\mathbb{Q}(\sqrt{3}): \mathbb{Q}]\). The connections between the remaining subgroups and the remaining intermediate fields are established similarly.

Here is a sample of the applications of Galois Theory. The following result is often used in arguments on field extensions:

**4.8 Proposition.** Let \(K \subseteq L\) be finite Galois and let \(K \subseteq M\) be an extension (suppose \(L\) and \(M\) are subfields of a larger field \(F\)). Then \(M \subseteq ML\) is finite Galois and \(G(ML/M)\) is isomorphic to a subgroup of \(G(L/K)\) (namely, with \(G(L/M \cap L)\)).
**Proof.** $L$ is the splitting field over $K$ of a separable polynomial $f \in K[X]$, so $ML$ is the splitting field over $M$ of $f$ (considered in $M[X]$). So, $M \subseteq ML$ is normal, finite and separable. Let $\sigma$ be an automorphism in $G := G(ML/M)$. Then $\sigma|_L \in G(L/K)$. Consider the group homomorphism $\text{res} : G \to G(L/K)$, $\text{res}(\sigma) = \sigma|_L$, $\forall \sigma \in L$. It is injective: $\forall \sigma \in G$ with $\sigma|_L = \text{id}$, we have $\sigma = \text{id}$ (since $\sigma$ and $\text{id}$ agree on $M$ and on $L$, they agree on $ML$). Thus $G \cong I$, where $I$ is the image of the res homomorphism; $I$ is a subgroup of $G(L/K)$. Let us show that $I = G(L/M \cap L)$. By the fundamental theorem, this is tantamount to the fact that the fixed fields of $I$ and $G(L/M \cap L)$ are equal. We have $L^I = \{ x \in L | \sigma(x) = x, \forall \sigma \in G(ML/M) \}$. But $\{ x \in ML | \sigma(x) = x, \forall \sigma \in G(ML/M) \} = M$, so $L^I = M \cap L$.

The classical Galois theory that we presented here using a “linearized” approach, due to Dedekind and Artin, has generalizations and counterparts in multiple directions:

**Infinite Galois Theory** treats the case of an extension that is algebraic, normal and separable, but not necessarily finite. The idea is to make the Galois group of the extension a *topological group* (by means of the Krull\(^9\) topology); the fundamental theorem reads in this case:

\(^9\)Wolfgang Adolf Ludwig Helmuth Krull (1899-1971), German mathematician.
The Galois connections establish bijections from the set of all intermediate fields of the extension to the set of all closed subgroups of the Galois group.

A Galois theory for commutative rings was developed by Chase, Harrison, Rosenberg, in a paper published in 1965.

Differential Galois theory looks into the problem of “explicit” solutions of differential equations. For an introduction and references, see for instance Gozard [1997].

Co-Galois theories. An example of such a theory is the Kummer theory that we present in the next chapter. The name reflects the fact that these theories establish inclusion preserving bijections between the lattice of intermediate fields and the lattice of subgroups of a certain group associated to the extension (as opposed to Galois theory, where they are order reversing). Recently, the Romanian mathematicians T. Albu and F. Nicolae devised a theory of this type that generalizes, among others, Kummer theory (T. Albu, Cogalois Theory, Marcel Dekker, New York, 2003).

Exercises

1. Let $L$ be the splitting field over $\mathbb{Q}$ of $X^3 - 2$. Find $G(L/\mathbb{Q})$ and all the subfields of $L$. (Hint: Look for the action of the automorphisms on the generators $\sqrt[3]{2}$ and $\omega$, where $\omega^2 + \omega + 1 = 0$).

2. The same problem, for the polynomial $X^4 - 2 \in \mathbb{Q}[X]$.

3. Let $G = G(L/\mathbb{Q})$, where $L = \mathbb{Q}(\sqrt[3]{2})$. Find $G$ and the fixed field of $G$.

4. Let $K$ be a field and let $K(X)$ be the rational function field over $K$. 
a) Let \( \psi : K(X) \to K(X) \) be the unique \( K \)-homomorphism with \( \psi(X) = X + 1 \). Determine the subgroup \( H \) generated by \( \psi \); find the fixed field of \( H \).

b) Prove that \( \varphi \in \text{Aut}_K(K(X)) = \text{Gal}(K(X)/K) \) iff there exist \( a, b, c, \) \( d \in K \), with \( ad - bc \neq 0 \) such that \( \varphi(X) = \frac{aX + b}{cX + d} \). In particular, \( H \neq \text{Gal}(K(X)/K) \), but these subgroups have the same fixed field. (Hint. Use IV.1.28c))

c) Show that \( \Phi : \text{GL}(2, K) \to \text{Aut}_K(K(X)), \ \Phi \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \varphi \), with \( \varphi(X) = \frac{aX + b}{cX + d} \), is a surjective group homomorphism, whose kernel is the subgroup \( S \) of scalar matrices (matrices of the type \( aI \), with \( a \in K \) and \( I \) is the identity matrix). The group \( \text{GL}(2, K)/S \) is called the projective linear group of degree 2 over \( K \), denoted \( \text{PGL}(2, K) \). So \( \text{Aut}_K(K(X)) \cong \text{PGL}(2, K) \).

5. Suppose \( L/K \) is a normal finite extension and \( S \) is the separable closure of \( K \) in \( L \). Show that:

a) \( K \subseteq S \) is normal and \( \text{res} : \text{Gal}(L/K) \to \text{Gal}(S/K) \), \( \text{res}(\sigma) = \sigma|_S \) is an isomorphism.

b) The fixed field of \( \text{Gal}(L/K) \) is \( I \), the purely inseparable closure of \( K \) in \( L \) (so \( I \subseteq L \) is Galois).

c) \( L = SI \).

6. Let \( K \subseteq L, K \subseteq E \) be Galois extensions such that \( L \cap E = K \). Then:
\[ \text{Gal}(L/E/K) \cong \text{Gal}(L/K) \times \text{Gal}(E/K). \]

7. Let \( L/K \) be a Galois extension and let \( E, F \) be intermediate fields. Show that:

a) \( \text{Gal}(L/EF) = \text{Gal}(L/E) \cap \text{Gal}(L/F) \).

b) \( \text{Gal}(L/E \cap F) = < \text{Gal}(L/E) \cup \text{Gal}(L/F) > \), the subgroup generated by \( \text{Gal}(L/E) \) and \( \text{Gal}(L/F) \).
VI. Applications of Galois Theory

VI.1 Ruler and compass constructions

Geometric ruler and compass constructions have a substantial historical interest. Such problems were one of the focal points of the antique Greek geometry. A number of celebrated problems stayed un-

---

10 A more general way to construct such extensions uses discriminants.
solved for millennia, despite the efforts of some of the best mathematicians. The solution (often in the negative) was given only in the 19\textsuperscript{th} century, at the dawn of the theory of field extensions. It is remarkable that the proof of impossibility of famous constructions requires only elementary concepts of field extensions theory, giving yet another demonstration of the power of the algebraic methods. The main difficulties seem to have been the transfer of the problem from Geometry to Algebra and a precise formulation of the concept of “constructible by ruler and compass”.

In this section, “construction” means exclusively “ruler and compass construction”.

We want to obtain general constructibility criteria, which will answer the following famous problems:

- “Angle trisection”: Construct an angle whose measure is 1/3 of the measure of a given angle.
- “Cube duplication”: Construct a cube whose volume is the double of the volume of a given cube.
- “Square the circle”: Construct a square whose area is equal to the area of a given circle.
- Construct a regular heptagon. More generally, for what natural numbers \( n \) is it possible to construct a regular polygon with \( n \) sides?

The \textit{rules} of construction are simple and well-known:

- A set of initial points is given (usually two points);
- For any two points (given, or already constructed as intersections of lines or circles), with the ruler one can draw the straight line passing through these points.
- The compass can draw the circle centered in a constructed (or given) point and passing through another constructed (or given) point.
These are the only constructions allowed. Note that drawing a line (a circle) does not mean that all points belonging to it are constructed. A point is constructed only if it is identified as an intersection between lines (or circles, or lines and circles).

Analyzing the various problems of construction, one realizes that all are equivalent to a problem of the following type:

“For a given set $S$ of points in the plane, construct a set $T$ of points (satisfying some property)”.

Indeed, the construction of a line reduces to the construction of two distinct points on that line; an angle is determined by three points (the vertex of the angle and a point on each side) etc. The reader is invited to formulate in this form the problems in the list above.

1.1 Definitions and notations. In what follows, $P$ denotes an Euclidian plane. If $A$ and $B$ are distinct points in $P$, $AB$ denotes the line determined by the points $A$ and $B$, $[AB]$ is the (closed) segment determined by $A$ and $B$, and $|AB|$ is the length of the segment $[AB]$. Let $S$ be a set of points in $P$ (called initially constructible points). Let $D_S$ be the set of the lines determined by two distinct points in $S$ and let $C_S$ be the set of circles centered in a point in $S$ and passing through a point in $S$.\(^{11}\)

The point $M \in P$ is called constructible\(^{12}\) in one step from $S$ if it satisfies one of the following conditions:

- $P$ belongs to the intersection of two distinct lines in $D_S$.

\(^{11}\) Sometimes $C_S$ is defined as the set of circles centered in a point in $S$ with radius equal to the distance between two arbitrary points in $S$. The definition we have adopted for $C_S$ corresponds to a “collapsible compass”: one cannot “transport” with the compass the distance between two points. The two definitions are in fact equivalent, in the sense that they lead to the same set of constructible points $C(S)$. Prove this!

\(^{12}\) We omit saying “by the ruler and compass” in what follows.
- $P$ belongs to the intersection of two distinct circles in $C_S$.
- $P$ belongs to the intersection of a circle in $C_S$ with a line in $D_S$.

Let $C_1(S)$ denote the set of points that are constructible in one step from $S$. Let $C_0(S) := S$ and define by recurrence $C_n(S) := C_1(C_{n-1}(S))$, $\forall n \in \mathbb{N}^*$.

A point in $\mathbb{P}$ is called constructible from $S$ if the point is in $C(S) := \bigcup_{n \in \mathbb{N}} C_n(S)$.

A line (a circle, an angle,…) is said to be constructible from $S$ if the line (the circle, the angle,…) is determined by points in $C(S)$.

If the set $S$ is understood, we say constructible instead of constructible from $S$.

1.2 Remark. a) $|S| = 1 \iff C_1(S) = S$. If one takes only one initially constructible point, nothing else can be constructed from it.

b) $S \subseteq C_1(S)$. Indeed, if $|S| \geq 2$, let $A, B \in S$. The point $A$ is the intersection of the line $AB$ with the circle of center $B$ passing through $A$. So, $C_n(S) \subseteq C_{n+1}(S)$, $\forall n \in \mathbb{N}$.

1.3 Lemma. Let $S \subseteq \mathbb{P}$ with $|S| \geq 2$ and let $A, B, C \in C(S)$ be distinct.

a) The symmetric of $B$ with respect to $A$ is constructible from $S$.

b) The midpoint $D$ of the segment $[AB]$ and the perpendicular on $[AB]$ in $D$ are constructible.

c) The perpendicular from $C$ on $AB$ is constructible.

d) If $C \notin AB$, then the parallel through $C$ to $AB$ is constructible.

Proof. Exercise in elementary geometry. \hfill $\square$

Assumptions and notations. We suppose from now on that $S$ is a subset of $\mathbb{P}$ with at least two points. Fix two distinct points $O$ and $I$ in $S$. There exists a unique system of Cartesian coordinates in the plane $\mathbb{P}$ such that $O$ is the origin, of coordinates $(0,0)$, and $I$ has coordinates $(1,0)$. The segment $OI$ has thus length 1. Let $\kappa : \mathbb{P} \to \mathbb{R} \times \mathbb{R}$ the mapping that associates to a point in $\mathbb{P}$ its coordinates in the system above.
We often identify a point \( M \in \mathbb{P} \) with the couple \((x, y)\) of its coordinates, and say “the point \((x, y)\)”. Thus, \( S \) can be seen as a subset of \( \mathbb{R} \times \mathbb{R} \). If \( T \) is a set of points in \( \mathbb{P} \), define \( T_R \) as the set of real numbers that occur as coordinates for the points in \( T \). Therefore,

\[
T_R := \{ x \in \mathbb{R} \mid \exists y \in \mathbb{R} \text{ such that } (x, y) \in T \text{ or } (y, x) \in T \}.
\]

In the hypotheses above, \( S_R \) contains at least 0 and 1.

As a consequence of the constructions in 1.3, the *coordinate axes Ox and Oy are constructible from O and I.*

**1.4 Definition.** The real number \( x \) is called *constructible from \( S \)* if the point of coordinates \((x, 0)\) is constructible from \( S \). If the set \( S \) is clear from the context, we say simply “\( x \) is constructible”.

In our hypotheses, 0 and 1 are always constructible.

**1.5 Proposition.** Let \( x \in \mathbb{R} \). The following assertions are equivalent:

\( a) \ x \) is constructible from \( S \).
\( b) \ The \ point \ (0, x) \) is constructible.
\( c) \ There \ exists \ y \in \mathbb{R} \ such \ that \ (x, y) \ is \ constructible. \)
\( d) \ There \ exists \ y \in \mathbb{R} \ such \ that \ (y, x) \ is \ constructible. \)

**Proof.** \( a) \Rightarrow b) \) Let \( M \) be the (constructible) point \((x, 0)\). By 1.3, construct the symmetric \( M' \) of \( M \) with respect to \( O \), then the midperpendicular of the segment \([MM']\) (i.e. the \( Oy \) axis). The point \((0, x)\) belongs to the intersection of the circle centered in \( O \) of radius \([OM]\) with \( Oy \).

The other implications are equally easy and are left to the reader. \( \square \)

The concept of constructible *complex* number is also natural (and useful, as we will see shortly):
1.6 Definition. The complex number $z = x + iy$ (where $x, y \in \mathbb{R}$) is called constructible from $S$ if the real numbers $x$ and $y$ are constructible from $S$.

The set $\mathbb{C}$ is in one-to-one correspondence with the plane $\mathbb{P}$ (composing the bijection $\kappa$ above with $(x, y) \mapsto x + iy$ from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{C}$). Let $S_C$ be the set of complex numbers that correspond to the points in $S$ (i.e., the affixes of the points in $S$):

$$S_C = \{x + iy \in \mathbb{C} | (x, y) \in S\}.$$ 

We have $\{0, 1\} \subseteq S_C$. It is natural to show that the following result holds:

1.7 Proposition. Let $a, b \in \mathbb{R}$. The following statements are equivalent:

a) The complex number $a + ib$ is constructible.

b) The point $(a, b)$ is constructible.

c) The real numbers $a$ and $b$ are constructible.

Proof. $a) \Rightarrow b)$ If $a + ib$ is constructible, then the points $A = (a, 0)$ and $B = (0, b)$ are constructible (use prop. 1.5). Thus, $(a, b)$ is the intersection of the perpendicular in $A$ on $Ox$ with the perpendicular in $B$ on $Oy$.

$b) \Rightarrow c)$ Let $(a, b)$ be a constructible point. Then the perpendiculars from $(a, b)$ on the axes $Ox$ and $Oy$ are constructible, so $(a, 0)$ and $(0, b)$ are constructible.

1.8 Remark. Let $\mathbb{K}$ be the set of constructible reals. The previous proposition says that the point $(x, y)$ is constructible if and only if $(x, y) \in \mathbb{K} \times \mathbb{K}$. In other words, $C(S) = \mathbb{K} \times \mathbb{K}$. The set of constructible complex numbers is $\mathbb{K}[i] = \{x + iy | x, y \in \mathbb{K}\}$.

\textsuperscript{13} Of course, we mean the real numbers constructible from $S$. Omitting $S$ from the notation $\mathbb{K}$ simplifies the notation.
1.9 Proposition. a) $\mathbb{K}$ is a subfield of $\mathbb{R}$, closed under square roots: for any $x \in \mathbb{K}$ with $x > 0$, $\sqrt{x} \in \mathbb{K}$. Moreover, $\mathbb{K}$ is the smallest subfield of $\mathbb{R}$ that includes $S_R$, the coordinates of the points in $S$, and is closed under square roots.

b) $\mathbb{K}[i]$ is a subfield of $\mathbb{C}$, closed under square roots and under complex conjugation: for any $z \in \mathbb{K}$, $\sqrt{z}$ and $\bar{z} \in \mathbb{K}$. Moreover, $\mathbb{K}[i]$ is the smallest subfield of $\mathbb{C}$ that includes the set $S_C$ of the affixes of points in $S$ and is closed under square roots and conjugation.

Proof. a) $0$ and $1$ belong to $\mathbb{K}$, since $O(0,0)$ and $I(0,1)$ are in $S$. Let $a, b \in \mathbb{K}$, $a, b \neq 0$. It is enough to prove that $-a, a + b, ab^{-1}$ are constructible. We use the constructions from 1.3.

The point $(a, 0)$ is constructible, so $(-a, 0)$, its symmetric with respect to $O$, is constructible, whence $-a \in \mathbb{K}$.

The constructible points $(0, a)$ and $(-a, 0)$ determine a line $d$. The perpendicular on $Ox$ in $(b, 0)$ is constructible and intersects $d$ in $(b, a + b)$. So, $a + b \in \mathbb{K}$.

For the constructibility of $ab^{-1}$, consider the line determined by $(b,0)$ and $(0, 1)$. The parallel to this line through $(a, 0)$ intersects $Oy$ in $(0, ab^{-1})$.

Suppose now that $a > 0$. Since $\mathbb{K}$ is a field, $((a + 1)/2, 0)$ is constructible. The circle centered in $((a + 1)/2, 0)$ passing through $(0, 0)$ intersects the perpendicular on $Ox$ in $(a, 0)$ in two points, of coordinates $(a, \sqrt{a})$ and $(a, -\sqrt{a})$. So, $\sqrt{a} \in \mathbb{K}$.

It is clear that $\mathbb{K}$ contains the coordinates of any point in $S$. The next lemma implies that, if $L$ is a subfield of $\mathbb{R}$, $L$ contains $S_R$ and $L$ is closed under square roots, then $\mathbb{K} \subseteq L$.

b) Since $\mathbb{K}$ is a subfield in $\mathbb{R}$, $\mathbb{K}[i]$ is a subfield in $\mathbb{C}$. If $a + bi$, with $a, b \in \mathbb{K}$, is an element in $\mathbb{K}[i]$, then its conjugate $a - bi \in \mathbb{K}[i]$. For proving that $\mathbb{K}[i]$ is closed under square roots, we use the trigonometric form of complex numbers. Let $z = r(\cos \alpha + isin \alpha)$, with $r > 0$, $r \cdot \cos \alpha$, $r \cdot \sin \alpha \in \mathbb{K}$. Then $\sqrt{z} = \pm \sqrt{r}(\cos(\alpha/2) + i \sin(\alpha/2))$. We have
$r^2 \cos^2 \alpha + r^2 \sin^2 \alpha = r^2$, so $r^2 \in \mathbb{K}$, $r$ and $\sqrt{r} \in \mathbb{K}$ (which is closed under square roots). We get also $\cos \alpha, \sin \alpha \in \mathbb{K}$. We have $\cos(\alpha/2) = \pm \sqrt{(1 + \cos \alpha)/2} \in \mathbb{K}$. Likewise, $\sin(\alpha/2) \in \mathbb{K}$, so $\sqrt{z} \in \mathbb{K}[i]$. The next lemma shows that any subfield of $\mathbb{C}$ containing the affixes of points in $S$ and is closed under conjugation and square roots must include $\mathbb{K}[i]$.

1.10 Lemma. a) Let $L$ be a subfield of $\mathbb{R}$ and let $(x, y) \in \mathbb{R} \times \mathbb{R}$. If $(x, y)$ is constructible in one step from $L \times L$, then $x, y \in L$ or there exists $u \in L, u > 0$, such that $x, y \in L(\sqrt{u})$.

b) If $L$ is a subfield of $\mathbb{R}$, closed under square roots, then $\mathbb{C}(L \times L) = L \times L$: the set of points constructible from $L \times L$ coincides with $L \times L$. In particular, if $S_R \subseteq L$, then $\mathbb{K} \subseteq L$.

c) Let $E$ be a subfield of $\mathbb{C}$ such that $E$ is closed under conjugation and square roots and $E \supseteq S_C$. Then $\mathbb{K}[i] \subseteq E$.

Proof. a) It is enough to prove that any point $M(x, y)$, constructible in one step from $L \times L$, is in $L \times L$ or in $L(\sqrt{u}) \times L(\sqrt{u})$ for some $u \in L, u > 0$. Use the notations in Definition 1.1.

Case I. $M(x, y)$ is the intersection of two lines in $D_L \times L$. Let $A_i(x_i, y_i) \in L \times L, \forall i \in \{1, 2, 3, 4\}$, such that $A_1A_2$ and $A_3A_4$ are not parallel and $M$ is their intersection. The equation of the line $A_1A_2$ is $(x - x_1)(y_2 - y_1) = (y - y_1)(x_2 - x_1)$. Since $L$ is a field, this equation can be written as $ax + by = c$, with $a, b, c \in L$. In the same way, the equation of $A_3A_4$ is $dx + ey = f$, for some $d, e, f \in L$. Then $M(x, y)$ is the solution of the system

$$\begin{align*}
ax + by &= c \\
dx + ey &= f
\end{align*}$$

The determinant of the system is nonzero, (otherwise the lines would be parallel); Cramer's formulas show that the solution $(x, y)$ of the system is in $L \times L$. 

VI.1 Ruler and compass constructions

Case II. M belongs to the intersection of a line in $D_L \times L$ with a circle in $C_L \times L$. The equation of a circle in $C_L \times L$ is $(x - a)^2 + (y - b)^2 = r^2$, for some $a, b, r \in L$. We saw that the equation of a line in $D_L \times L$ is $dx + ey = f$, for some $d, e, f \in L$. Then $M(x, y)$ is the solution of the system:

$$\begin{cases}
(x-a)^2 + (y-b)^2 = r^2 \\
 dx + ey = f
\end{cases}$$

Suppose $d \neq 0$ (if $d = 0$, then $e \neq 0$ and swap $d$ with $e$). So, $x = -d^{-1}ey - d^{-1}f$. The first equation becomes $py^2 + qy + s = 0$, for some $p, q, s \in L$. This equations must have real solutions (the circle and the line intersect), $y = 2p^{-1} \cdot (-q \pm \sqrt{q^2 - 2sp})$. So, $x, y \in L(\sqrt{u})$, where $u = q^2 - 2sp \in L$.

Case III. M belongs to the intersection of two circles in $C_L \times L$. We must show that the solutions (if any) of a system of the form

$$\begin{cases}
(x-a)^2 + (y-b)^2 = r^2 \\
x-c)^2 + (y-d)^2 = t^2
\end{cases}$$

where $a, b, r, c, d, t \in L$, are in $L(\sqrt{u}) \times L(\sqrt{u})$ for some $u \in L, u > 0$. Subtracting the equations, the quadratic terms cancel out and we obtain linear equation in $x$ and $y$; together with one of the initial equations, we obtain a system of the type studied at case II.

b) The subfield $L$ being closed under square roots, the result follows from a).

If $S_R \subseteq L$ (i.e. $S \subseteq L \times L$), then $C(S) \subseteq C(L \times L)$. By a), $C(L \times L) = L \times L$. Thus, $K \subseteq L$.

c) Let $L = E \cap \mathbb{R}$. If $x \in L, x > 0$, then $\sqrt{x} \in E$ ($E$ is closed under square roots) and $\sqrt{x} \in \mathbb{R}$, so $\sqrt{x} \in L$. Thus, $L$ is a subfield of $\mathbb{R}$, closed under square roots. On the other hand, if $x, y \in \mathbb{R}$ and $x + iy \in E$, then $x - iy \in E$ ($E$ is closed under conjugation), so $x, y \in E$. Since $x, y \in \mathbb{R}, x, y \in L$. Conversely, if $x, y \in L$, then $x + iy \in E$, because $i = \sqrt{-1} \in E$. So, $E = L[i] = \{x + iy \mid x, y \in L\}$. 
The hypothesis $S_C \subseteq E$ is thus equivalent to $S_R \subseteq L$; by $a)$, we have $K \subseteq L$, so $K[i] \subseteq L[i] = E$. \hfill \square

We have proven that the geometric statement “the point $M(x, y)$ is constructible from the set of points $S$” has a purely algebraic form: “$x$ and $y$ belong to the smallest subfield of $\mathbb{R}$ that includes $\mathbb{Q}(S_R)$ and is closed under square roots”. This algebraic translation allows us to say about some real number $x$: “$x$ is constructible from $U$” (where $U$ is a set of real numbers), meaning that “$x$ belongs to the smallest subfield of $\mathbb{R}$ that includes $\mathbb{Q}(U)$ and is closed under square roots”.

Note also that any rational number is constructible (from any $U \subseteq \mathbb{R}$, $|U| \geq 2$).

**Theorem.** A real number $x$ is constructible from $S$ if and only if there exists a tower $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ of subfields of $\mathbb{R}$, such that $x \in K_n$, $K_0 = \mathbb{Q}(S_R)$ and $[K_i : K_{i-1}] \leq 2$, for any $i$, $1 \leq i \leq n$.

**Proof.** “$\Rightarrow$” It is enough to prove by induction on $m$ that, for any finite set of points $T$, if $T \subseteq C_m(S)$, then there exists a tower of extensions $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ such that $T_R \subseteq K_n$ and $[K_i : K_{i-1}] \leq 2$, $\forall i \in \{1, \ldots, n\}$. If $m = 0$, $T \subseteq S$, so $T_R \subseteq S_R \subseteq \mathbb{Q}(S_R) = K_0$. Let $m > 0$ and let $T = \{A_1, \ldots, A_r\}$ be a finite subset of $C_m(S)$. This means that there exists a finite set $U \subseteq C_{m-1}(S)$ such that any point in $T$ is constructible in one step from $U$. From the induction hypothesis, there is a tower $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ such that $U_R \subseteq K_n$ and $[K_i : K_{i-1}] \leq 2$, $\forall i \in \{1, \ldots, n\}$. For any point $A_s \in T$ ($s \in \{1, \ldots, r\}$), lemma 1.10. $a)$ says that $A_s$ has its coordinates in $K_n$ (then set $u_s = 0$) or in a quadratic extension of $K_n$ of the form $K_n(\sqrt{u_s})$, for some $u_s \in K_n$, $u_s > 0$. So, $T_R \subseteq K_n(\sqrt{u_1}, \ldots, \sqrt{u_r})$ and we have a tower of extensions.

---

$^1$The result belongs to M.L. Wantzel, who proved it and published it in 1837 (when he was “élève-ingénieur des Ponts-et-Chaussées”). It seems though that Gauss knew as early as 1796 this criterion of constructibility.
VI.1 Ruler and compass constructions

$K_0 \subseteq \ldots \subseteq K_n \subseteq K_n(\sqrt{u_1}) \subseteq \ldots \subseteq K_n(\sqrt{u_1},\ldots,\sqrt{u_r})$, satisfying the condition that each extension has degree at most 2.

“$\subseteq$” We prove by induction on $n$ that, for any tower $\mathbb{Q}(S_R) = K_0 \subseteq \ldots \subseteq K_n \subseteq \mathbb{R}$, with $[K_i : K_{i-1}] \leq 2, 1 \leq i \leq n$, we have $K_n \subseteq \mathbb{K}$ (recall $\mathbb{K}$ is the field of the real numbers constructible from $S$). If $n = 0$, then $K_0 = \mathbb{Q}(S_R)$. Using 1.9.a), we deduce that $\mathbb{K}$ includes the subfield of $\mathbb{R}$ generated by $S_R$, which is exactly $\mathbb{Q}(S_R)$. If $n > 0$, $[K_n : K_{n-1}] \leq 2$ implies either that $K_n = K_{n-1}$ (and the induction hypothesis shows that we are done) or that $K_n$ is an extension of degree 2 of $K_{n-1}$. In this case there exists $u \in K_{n-1}$, $u > 0$, such that $K_n = K_{n-1}(\sqrt{u})$. By induction, we know that $K_{n-1} \subseteq \mathbb{K}$; since $\mathbb{K}$ is closed under square roots, we deduce $K_n \subseteq \mathbb{K}$. $\square$

1.12 Corollary. (Necessary condition for constructibility) Let $x$ be a real number constructible over $S$ and let $K_0 = \mathbb{Q}(S_R)$. Then $x$ is algebraic over $K_0$ and its degree over $K_0$ is a power of 2 (there exists $e \in \mathbb{N}$ such that $[K_0(x) : K_0] = 2^e$).

Proof. The extension $K_n$ in the previous theorem is such that $x \in K_n$ and $[K_n : K_0]$ is a power of 2. Thus, $x$ is algebraic over $K_0$ and $K_0 \subseteq K_0(x) \subseteq K_n$; so, $[K_0(x) : K_0]$ divides $[K_n : K_0]$, which means that $[K_0(x) : K_0]$ is a power of 2. $\square$

We can easily prove now that some classic ruler and compass constructions problems are impossible:

1.13 Proposition. a) The trisection of an angle of measure $\theta$ is equivalent to the construction of the point $(\cos(\theta/3), \sin(\theta/3))$ from $\{O,$
I, \( (\cos \theta, \sin \theta) \). A 60° angle cannot be trisected by ruler and compass (a 20° angle is not constructible).

b) Cube duplication is impossible (it is impossible to construct a cube whose volume is the double of the volume of a given cube).

c) Squaring the circle is impossible (it is impossible to construct a square whose area is equal to the area of a given circle).

**Proof.** a) Constructing an angle of measure \( \alpha \) is equivalent to constructing the point \( (\cos \alpha, \sin \alpha) \). Indeed, if the angle has its vertex in \( O \) one of the sides is \( Ox \), the intersection of the other side with the unit circle is \( (\cos \alpha, \sin \alpha) \) or \( (\cos \alpha, -\sin \alpha) \). Conversely, if \( P(\cos \alpha, \sin \alpha) \) is given, then the angle between \( OP \) and \( Ox \) has measure \( \alpha \).

On the other hand, if \( \cos \alpha \) is constructible, then \( \sin \alpha = \pm \sqrt{1 - \cos^2 \alpha} \) is constructible. Thus, trisecting the angle of measure \( \theta \) is tantamount to the constructibility of \( \cos(\theta/3) \) from \( \{\cos \theta\} \). The formula \( \cos(3x) = 4\cos^3 x - 3\cos x \) implies that \( u := \cos(\theta/3) \) satisfies the equation:

\[
4u^3 - 3u = \cos \theta.
\]

We are led to the study of the polynomial \( 4X^3 - 3X - \cos \theta \), with coefficients in \( \mathbb{Q}(\cos \theta) \). For \( \theta = 60^\circ \), we obtain that \( u := \cos 20^\circ \) is a root of \( g := 8X^3 - 6X - 1 \in \mathbb{Q}[X] \). We have \( \text{Irr}(u, \mathbb{Q}) = g \) (since \( g \) has no rational roots), so \( [\mathbb{Q}(u) : \mathbb{Q}] = 3 \). This shows that \( u \) is not constructible, since \( [\mathbb{Q}(u) : \mathbb{Q}] \) is not a power of 2.

b) Choose the unit length to be the length of the side of the cube. Thus, the initially constructible points are \( O \) and \( I \). The cube having double the volume has the side \( \sqrt[3]{2} \). The real number \( \sqrt[3]{2} \) is not constructible, since \( [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3 \) is not a power of 2.

---

\( ^{15} \) This result does not say that no angle can be trisected (for instance, a 90° angle can be trisected), but that some angles (the 60° angle) cannot be trisected. Thus, there exists no ruler and compass construction of the trisection of an arbitrary angle.
VI.1 Ruler and compass constructions

285

c) Choosing the unit length to be the radius of the circle, the initially constructible points are $O$ and $I$. The area of the circle of radius 1 is $\pi$, so the side of the square with area $\pi$ is $\sqrt{\pi}$. But $\pi$ is transcendental over $\mathbb{Q}$ and $\sqrt{\pi}$ is also transcendental, so it is not constructible.

For the formulation of a necessary and sufficient criterion of constructibility, the following complex version of Theorem 1.11 is useful:

1.14 Theorem. a) The complex number $z$ is constructible from $S$ (i.e. $z \in \mathbb{K}(i)$) if and only if there exists a chain of subfields of $\mathbb{C}$, $L_0 \subseteq L_1 \subseteq \ldots \subseteq L_n$, such that $z \in L_n$, $L_0 = \mathbb{Q}(S_R)(i)$ and $[L_t : L_{t-1}] \leq 2$, for any $t \in \{1, \ldots, n\}$.

b) If $z$ is a complex number constructible over $S$, then $z$ is algebraic over $L_0 = \mathbb{Q}(S_R)(i)$ and its degree over $L_0$ is a power of 2 (there exists $e \in \mathbb{N}$ such that $[L_0(x) : L_0] = 2^e$).

Proof. a) “$\Rightarrow$” Let $z = x + yi$, with $x, y \in \mathbb{K}$. By 1.11, there exists a chain $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ of subfields of $\mathbb{R}$, with $K_0 = \mathbb{Q}(S_R)$, $x, y \in K_n$ and $[K_t : K_{t-1}] \leq 2$, for any $t \in \{1, \ldots, n\}$. Let $L_t := K_t(i)$. The chain of subfields $L_0 \subseteq L_1 \subseteq \ldots \subseteq L_n$ satisfies the required conditions.

“The same technique as in Theorem 1.11 is used. By induction on $n$, we show that for any chain $\mathbb{Q}(S_R)(i) = L_0 \subseteq L_1 \subseteq \ldots \subseteq L_n$ of subfields of $\mathbb{C}$, with $[L_t : L_{t-1}] \leq 2$, $\forall t \in \{1, \ldots, n\}$, we have $L_n \subseteq \mathbb{K}(i)$. If $n = 0$, apply 1.9. For $n > 0$, $[L_n : L_{n-1}] \leq 2$ implies $L_n = L_{n-1}$ or $L_n$ is a quadratic extension of $L_{n-1}$. Since $\mathbb{K}(i)$ is closed under square roots, and $L_n \subseteq \mathbb{K}(i)$ by hypothesis, $L_n \subseteq \mathbb{K}(i)$.

b) Exercise.

1.15 Theorem. (Characterization of constructible real numbers) Suppose $U \subseteq \mathbb{R}$, $x \in \mathbb{R}$ is algebraic over $\mathbb{Q}(U)$ and $N$ is the normal closure of $\mathbb{Q}(U) \subseteq \mathbb{Q}(U)(x)$. Then $x$ is constructible from $U$ if and only if $[N : \mathbb{Q}(U)]$ is a power of 2.
Proof. Let \( g := \text{Irr}(x, \mathbb{Q}(U)) \in \mathbb{Q}(U)[X] \). Then \( N \) is the splitting field of \( g \) over \( \mathbb{Q}(U) := K_0 \). We must show that \( x \in \mathbb{K} \) (the numbers constructible from \( U \)) if and only if \( [N : K_0] \) is a power of 2.

\[ \Rightarrow \] Theorem 1.11 shows there exists a tower \( K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n \) of subfields of \( \mathbb{R} \), such that \([K_t : K_{t-1}] \leq 2\), \( \forall t \in \{1, \ldots, n\} \) and \( x \in K_n \). Let \( E \) be the normal closure of \( K_0 \subseteq K_n \). Since \( x \in E \) and \( E \) is normal over \( K_0 \), \( g \) splits over \( E \), so \( N \subseteq E \). On the other hand, \( E = K_0 \bigcup \{ \sigma(K_n) \mid \sigma \in H \} \), where \( H \) is the set of the \( K_0 \)-homomorphisms from \( K_n \) to \( \mathbb{C} \) (see 2.7). For any \( \sigma \in H \), the chain of extensions \( K_0 = \sigma(K_0) \subseteq \sigma(K_1) \subseteq \ldots \subseteq \sigma(K_n) \) has the property that \([\sigma(K_t) : \sigma(K_{t-1})] \leq 2\), \( \forall t \in \{1, \ldots, n\} \). The proof of theorem 1.14 implies that \( \sigma(K_n) \subseteq \mathbb{K}(i) \). So, \( E \subseteq \mathbb{K}(i) \) and thus \( N \subseteq \mathbb{K}(i) \). Let \( \alpha \) be a primitive element of the separable finite extension \( K_0 \subseteq N \). Because \( \alpha \in \mathbb{K}(i) \), theorem 1.14.b) guarantees that \([K_0(\alpha) : K] = [N : K_0] \) is a power of 2.

\[ \Leftarrow \] The Galois extension \( K_0 \subseteq N \) has the degree \( 2^e \) for some \( e \), so its Galois group \( G \) has \( 2^e \) elements. From (Appendix 6, Proposition 11) there exists a chain of subgroups of \( G \), \( G = G_0 \supseteq G_1 \supseteq \ldots \supseteq G_e = \{\text{id}\} \), with \([G_{t-1} : G_t] = 2\), \( \forall t \in \{1, \ldots, e\} \) (so \(|G_t| = 2^{e-t}\)). Denoting by \( K_t \) the fixed field of \( G_t \), \( \forall t \in \{1, \ldots, e\} \), we have \( K_0 \subseteq K_1 \subseteq \ldots \subseteq K_e = N \) (we applied the Galois correspondence). We have \([K_t : K_{t-1}] = [G_{t-1} : G_t] = 2\), \( \forall t \in \{1, \ldots, n\} \). Theorem 1.14. implies that \( N \subseteq \mathbb{K}(i) \). Therefore \( x \in \mathbb{R} \cap \mathbb{K}(i) = \mathbb{K} \).

1.16 Remark. Keeping the same notations, we also have: \( x \) is constructible from \( U \) if and only if there exists a normal extension \( L \) of \( \mathbb{Q}(U) \), with \([L : \mathbb{Q}(U)] \) a power of 2 and \( x \in L \). Indeed, if \( x \) belongs to such an extension \( L \), then the normal closure of \( \mathbb{Q}(U) \subseteq \mathbb{Q}(U)(x) \) is included in \( L \) and has its degree over \( \mathbb{Q}(U) \) a power of 2.

We now solve the classical problem of the constructibility of regular polygons.
1.17 Theorem. (Gauss, 1801) Let \( n \geq 3 \). The regular \( n \)-gon is constructible by ruler and compass if and only if the prime factor decomposition of \( n \) is \( n = 2^e \cdot p_1 \cdot ... \cdot p_t \), where \( e, t \in \mathbb{N} \) and \( p_1, ..., p_t \) are distinct primes of the form \( 2^{2^m} + 1 \), for some \( m \in \mathbb{N} \).\(^{16}\)

Proof. Here, “constructible” means “constructible from two points (\( O \) and \( I \))”. To construct a regular \( n \)-gon is equivalent to construct the angle \( \frac{2\pi}{n} \). Let \( a, b \in \mathbb{N}^* \), with \( (a, b) = 1 \). We claim that the angle \( \frac{2\pi}{ab} \) is constructible if and only if the angles \( \frac{2\pi}{a} \) and \( \frac{2\pi}{b} \) are constructible. The direct implication is simple: for a given angle (for instance \( \frac{2\pi}{ab} \)), any multiple of it can be constructed (so the angles \( \frac{2\pi}{a}, \frac{2\pi}{b} \) are constructible). Conversely, let \( u, v \in \mathbb{Z} \) such that \( ua + vb = 1 \). Multiplying by \( \frac{2\pi}{ab} \), we get \( u \cdot 2\pi/b + v \cdot 2\pi/a = 2\pi/ab \) and we apply the fact that if the angles \( \alpha \) and \( \beta \) are constructible, then, for any \( u, v \in \mathbb{Z} \), the angle \( u\alpha + v\beta \) is constructible.

Thus, if \( n = p_1^{\alpha_1} \cdot ... \cdot p_k^{\alpha_k} \), with \( p_1, ..., p_k \) distinct primes, the construction of the angle \( \frac{2\pi}{n} \) is equivalent to the construction of every angle \( \frac{2\pi}{p_i^{\alpha_i}} \), for any \( i \in \{1, ..., k\} \). This allows us to suppose from now on that \( n \) is a power of a prime.

If \( n = 2^e, e > 1 \), then the angle \( \frac{2\pi}{n} \) is constructible (induction on \( e \): the angle \( \frac{2\pi}{2^e} \) is obtained by bisecting the angle \( \frac{2\pi}{2^{e-1}} \)).

At 1.13. we saw that the angle \( \alpha \) is constructible if and only if \( \cos \alpha \) is constructible.

---

\(^{16}\) Such a prime number is called a Fermat prime. For \( m = 0, 1, 2, 3, 4 \), one obtains 3, 5, 17, 257, 65537, which are indeed prime. No other Fermat primes have been found yet. Using computers, it has been proven that the above are all the Fermat primes less than \( 10^{15000} \).

\(^{17}\) We say “the angle \( u \)” instead of “the angle of measure \( u \)”.
Suppose from now on that \( n = p^\alpha, \) \( p \) a prime, \( p \neq 2, \) \( \alpha \in \mathbb{N}^* \). We prove that \( \cos(2\pi/p^\alpha) \) is constructible \( \iff \alpha = 1 \) and \( p \) is a Fermat prime.

Suppose \( \cos(2\pi/n) \) is constructible. Then, using 1.12, \([\mathbb{Q}(\cos(2\pi/n)) : \mathbb{Q}]\) is a power of 2. Let \( \zeta := \cos(2\pi/n) + i\sin(2\pi/n) \). The next lemma says that \([\mathbb{Q}(\zeta) : \mathbb{Q}] = 2\cdot[\mathbb{Q}(\cos(2\pi/n)) : \mathbb{Q}]\). On the other hand, the degree of the cyclotomic extension \( \mathbb{Q} \subseteq \mathbb{Q} (\zeta) \) is \( \varphi(n) = \varphi(p^\alpha) = p^\alpha - p^{\alpha-1} \), so \( p^\alpha - p^{\alpha-1} \) is a power of 2. Let \( e \in \mathbb{N} \) with \( p^{\alpha-1}(p - 1) = 2^e \). If \( \alpha \geq 2 \), then \( p \mid 2^e \), absurd. So \( \alpha = 1 \) and \( p - 1 = 2^e \), whence \( p = 2^e + 1 \). Suppose \( e \) is not a power of 2. Then \( e = a \cdot 2^b \), for some \( a, \ b \in \mathbb{N}, \ a \geq 3, \ a \ odd \). So, \( p = 2^b + 1 \). Let \( d := 2^b \), so \( p = d^a + 1 \). Since \( a \) is odd , \( d^a + 1 \) is divisible by \( d + 1 \), so \( p \) is not a prime, contradiction. Hence \( e \) must be a power of 2, which means that \( p \) is a Fermat prime.

Conversely, let \( p \) be a Fermat prime. Let us prove that \( \cos(2\pi/p) \) is constructible. The extension \( \mathbb{Q} \subseteq \mathbb{Q} (\zeta) \) is normal, of degree \( \varphi(p) = p - 1 \), which is a power of 2. Since \( \zeta = \cos(2\pi/p) + i\sin(2\pi/p) \) and \( \zeta^{-1} = \cos(2\pi/p) - i\sin(2\pi/p) \) are in \( \mathbb{Q} (\zeta) \), \( \cos(2\pi/p) \) belongs to \( \mathbb{Q} (\zeta) \). The remark following Theorem 1.15. applies and we obtain the constructibility of \( \cos(2\pi/p) \).

1.18 Lemma. Let \( n \in \mathbb{N}, \ n \geq 2 \) and let \( \zeta \) be a primitive complex \( n \)th root of unity. Then

\[ [\mathbb{Q}(\zeta) : \mathbb{Q}(\cos(2\pi/n))] = 2. \]

Proof. We may suppose that \( \zeta = \cos(2\pi/n) + i\sin(2\pi/n) \). Clearly, \( \zeta \not\in \mathbb{Q}(\cos(2\pi/n)) \), so \([\mathbb{Q}(\zeta) : \mathbb{Q}(\cos(2\pi/n))] \geq 2 \). But \( i\sin(2\pi/n) \) is a root of \( \cos(2\pi/n)^2 - X^2 = 1 \), so \([\mathbb{Q}(\zeta) : \mathbb{Q}(\cos(2\pi/n))] \leq 2. \)

1.19 Example. A regular heptagon is impossible to construct by ruler and compass, as is a regular 9-gon. In fact, for \( n \leq 20 \), a regular \( n \)-gon is constructible \( \iff n \in \{3, 4, 5, 6, 8, 10, 12, 15, 16, 17\} \).
VI.2 Trace and norm

If \( K \subseteq L \) is a finite extension and \( x \in L \), the conjugates of \( x \) over \( K \) are the roots of the minimal polynomial of \( x \) over \( K \). The sum (and the product) of these conjugates, each occurring with its multiplicity, are elements in \( K \), as the Viète relations show. For instance, if \( x = a + b \sqrt{d} \), with \( d \) a square free integer and \( a, b \in \mathbb{Q} \), then \( a - b \sqrt{d} \) is the only conjugate of \( x \) over \( \mathbb{Q} \). The sum of all conjugates of \( a + b \sqrt{d} \) is thus \( 2a \), and the product of all its conjugates is \( a^2 - db^2 \). This is what we called “the norm of \( a + b \sqrt{d} \)” and appeared in the study or arithmetic properties of the ring \( \mathbb{Z}[\sqrt{d}] \). These are particular cases of concepts that we study now.

Although we use the notions of trace and norm mainly in the case of a finite extension of fields, the definitions and the proofs are the same as in the more general case of an extension of commutative rings \( R \subseteq S \) such that \( S \) is a free \( R \)-module of finite rank. The reader less familiar with these concepts may assume in what follows that \( R \subseteq S \) is a finite extension of fields (so, \( S \) is an \( R \)-vector space of finite dimension). In this case, “free module” translates into “vector space”, “module homomorphism” means “linear mapping” etc.

Let \( R \subseteq S \) be an extension of commutative rings with identity\(^{18}\), such that \( S \), with its canonical \( R \)-module structure, is free and has rank \( n \). We recall some facts on free modules from II.4. We denote \( \text{End}_R(S) := \{ \varphi : S \to S \mid \varphi \text{ is an } R\text{-module homomorphism} \} \). \( (\text{End}_R(S), +, \circ) \) is a ring with identity; "+" denotes homomorphism addition and "\( \circ \)" the usual map composition. Let \( e = (e_1, \ldots, e_n) \) be a basis of the free \( R \)-module \( S \). If \( \varphi \in \text{End}_R(S) \), the matrix of \( \varphi \) in the basis \( e \) is the \( n \times n \) matrix \( M_e(\varphi) = (a_{ij})_{1 \leq i, j \leq n} \), where \( a_{ij} \in R \) are defined by the relations

\(^{18}\) In other words, \( S \) is a commutative ring with identity and \( R \) is a subring of \( S \) containing the identity.
\[ \varphi(e_i) = \sum_{j=1}^{n} a_{ij}e_j, \quad \forall i \in \{1, \ldots, n\}. \]

We obtain a function \( M_e : \text{End}_R(S) \rightarrow M_n(R), \varphi \mapsto M_e(\varphi) \), which is a ring anti-isomorphism. If \( \mathbf{f} = (f_1, \ldots, f_n) \) is an \( n \)-uple of elements in \( S \), then there exists a unique matrix \( U = (u_{ij})_{1 \leq i, j \leq n} \in M_n(R) \) such that:

\[ f_i = \sum_{j=1}^{n} u_{ij}e_j. \]

\( \mathbf{f} = (f_1, \ldots, f_n) \) is a basis in \( R \cdot S \iff U \text{ is invertible in } M_n(R) \iff \det U \in U(R) \) (the group of units of \( R \)); in this case,

\[ M_f(\varphi) = U \cdot M_e(\varphi) \cdot U^{-1}. \]

The characteristic polynomial of a vector space endomorphism inspires the following definition (cf. III.4.18.):

**2.1 Definition.** Let \( I \in M_n(R) \) the identity matrix and let \( A = (a_{ij}) \in M_n(R) \). The matrix:

\[
XI - A := \begin{bmatrix}
X - a_{11} & -a_{12} & \cdots & -a_{1n} \\
-a_{21} & X - a_{22} & \cdots & -a_{2n} \\
\vdots & & \ddots & \vdots \\
-a_{n1} & -a_{n2} & \cdots & X - a_{nn}
\end{bmatrix} \in M_n(R[X])
\]

is called the **characteristic matrix of** \( A \). Let \( \varphi \in \text{End}_R(S) \) and let \( \mathbf{e} \) be an arbitrary basis in \( R \cdot S \). Define

\[ P(\varphi) := \det(XI - M_e(\varphi)) \in R[X], \]

called the **characteristic polynomial** of \( \varphi \).

The characteristic polynomial of \( \varphi \) is correctly defined (it is independent of the choice of a basis): if \( \mathbf{f} \) is another basis, there exists \( U \in U(M_n(R)) \) such that \( M_f(\varphi) = U \cdot M_e(\varphi) \cdot U^{-1} \), so \( XI - M_f(\varphi) = U \cdot (XI - M_e(\varphi)) \cdot U^{-1} \). Thus,

\[ \det(XI - M_f(\varphi)) = \det U \cdot \det(XI - M_e(\varphi)) \cdot \det(U^{-1}) = \det(XI - M_e(\varphi)). \]
Let \( P(\varphi) := X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0; \) the coefficients \( a_i \in R \) depend only on \( \varphi \) (not on the choice of the basis), so the following definitions are correct:

\[
\begin{align*}
\text{Tr}(\varphi) &:= -a_{n-1} \quad (\text{called the trace of } \varphi) \\
\det(\varphi) &:= (-1)^n a_0 \quad (\text{called the determinant of } \varphi).
\end{align*}
\]

If \( A = (a_{ij}) \in M_n(R) \) is the matrix of \( \varphi \) in some basis, then

\[
\begin{align*}
\text{Tr}(\varphi) &= \text{Tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}; \\
\det(\varphi) &= \det(A).
\end{align*}
\]

For any \( x \in S \), the function

\[
\theta_x : S \to S, \; \theta_x(y) = xy, \; \forall y \in S,
\]

is an \( R \)-module homomorphism. We apply the definitions above to \( \theta_x \in \text{End}_R(S) \). The following notations and terms are used:

\[
\begin{align*}
P(\theta_x) &= P(x, S/R) \in R[X] \text{ is called the characteristic polynomial of } x; \\
\text{Tr}(\theta_x) &= \text{Tr}_{S/R}(x) \in R \text{ is called the trace of } x; \\
\det(\theta_x) &= \text{N}_{S/R}(x) \in R \text{ is called the norm of } x.
\end{align*}
\]

It is clear that these notions depend not only on \( x \), but on the extension \( S \) of \( R \). The notations we use take this into account, but not the terminology, so some caution is recommended. We defined therefore the mappings \( \text{trace } \text{Tr}_{S/R} : S \to R \) and \( \text{norm } \text{N}_{S/R} : S \to R \).

### 2.2 Remark.

\( \theta : (S, +, \cdot) \to (\text{End}_R(S), +, \circ), \; x \mapsto \theta_x \), is an \( R \)-module homomorphism and also a ring homomorphism (in other words, \( \theta \) is an \( R \)-algebra homomorphism). Prove this!

Using this remark, the next properties are easy to prove:

### 2.3 Proposition.

In the conditions above, we have, \( \forall x, y \in S, \; \forall a \in R: \)

\[
\begin{align*}
\text{Tr}_{S/R}(x + y) &= \text{Tr}_{S/R}(x) + \text{Tr}_{S/R}(y); \\
\text{N}_{S/R}(xy) &= \text{N}_{S/R}(x) \cdot \text{N}_{S/R}(y); \\
\text{Tr}_{S/R}(ax) &= a \text{Tr}_{S/R}(x); \\
\text{N}_{S/R}(ax) &= a^n\text{N}_{S/R}(x); \\
\text{Tr}_{S/R}(a) &= na; \\
\text{N}_{S/R}(a) &= a^n.
\end{align*}
\]

So, \( \text{Tr}_{S/R} : S \to R \) is an \( R \)-module homomorphism.
The case we are interested in is when \( R \) and \( S \) are fields. Fix a finite extension of fields \( K \subseteq L \). We suppose that all algebraic extensions of \( K \) are subfields of \( \Omega \), a fixed algebraic closure of \( K \).

### 2.4 Theorem

Let \( K \subseteq L \) be a finite extension of fields, let \( x \in L \) and let \( P(x, L/K) \) be its characteristic polynomial. Let \( H := \text{Hom}_K(L, \Omega) \) be the set of \( K \)-homomorphisms from \( L \) to \( \Omega \). Then:

\[
P(x, L/K) = \text{Irr}(x, K)^{[L : K(x)]}
\]

(1)

\[
\text{Irr}(x, K) = P(x, K(x)/K) \quad \text{and} \quad P(x, L/K) = \left(\prod_{\sigma \in H} (X - \sigma(x))\right)^d,
\]

(2)

where \( d = [L : K] \) is the inseparability degree of \( L \) over \( K \). In particular:

\[
\text{Tr}_{L/K}(x) = [L : K(x)] \cdot \text{Tr}_{K(x)/K}(x) = d \cdot \left(\sum_{\sigma \in H} \sigma(x)\right)
\]

(3)

\[
N_{L/K}(x) = (N_{K(x)/K(x)})^{[L : K(x)]} = \left(\prod_{\sigma \in H} \sigma(x)\right)^d.
\]

(4)

**Proof.** Note that, for any \( g \in K[X] \), \( g(\theta) = \theta_{g(x)} \) (use 2.2 for the proof). Let \( P := P(x, L/K) \in K[X] \). The Cayley-Hamilton theorem for \( \theta \in \text{End}_K(L) \) says that \( P(\theta) = 0 \). So, \( \theta_{P(x)} = 0 \), which means \( P(x) = \theta_{P(x)}(1) = 0 \).

Let \( f := \text{Irr}(x, K) \). We have \( f \mid P \), since \( P(x) = 0 \). Also, \( f(\theta) = \theta_{f(x)} = 0 \); if \( g \in K[X] \) is such that \( g(\theta) = 0 \), then \( \theta_{g(x)} = 0 \), so \( g(x) = 0 \), which means \( f \mid g \). This shows that \( f \) is the minimal polynomial of the endomorphism \( \theta \). The Frobenius theorem implies that \( P \) and \( f \) have the same irreducible factors. Since \( f \) is irreducible, \( f \) is the only irreducible factor of \( P \). Thus, \( P = f^t \), for some \( t \in \mathbb{N}^* \). Taking the degrees, we obtain \( n := \deg P = [L : K] = t \cdot \deg f = t \cdot [K(x) : K] \), so \( t = [L : K(x)] \). This proves (1).

Let \( Q := P(x, K(x)/K) \). As for \( P(x, L/K) \), we have \( Q(x) = 0 \), so \( f \mid Q' \); since \( \deg Q = [K(x) : K] = \deg f \), we obtain \( f = Q \). Identifying the coefficients of \( X^{n-1} \) in (1), written as \( P(x, L/K) = P(x, K(x)/K)^{[L : K(x)]} \),
we get \( \text{Tr}_{L/K}(x) = [L : K(x)] \cdot \text{Tr}_{K(x)/K}(x) \), which is the first equality in (3).

Setting \( X = 0 \) in (1), we obtain the first equality in (4).

Formula (1) says that \( \text{Irr}(x, K) = P(x, K(x)/K) \) and reduces the computation of \( P(x, L/K) \) to the case \( L = K(x) \). Assume for the moment that \( L = K(x) \). Let \( n := |H| = [L : K] \), \( m \) (the separable degree of \( L \) over \( K \)). The polynomial \( f \) has \( n \) distinct roots in \( \Omega \), namely \( \{ \sigma x | \sigma \in H \} \). If \( m \) is the multiplicity of the root \( x \) of \( f \) and \( \sigma \in H \), then the root \( \sigma x \in \Omega \) has also multiplicity \( m \). Indeed, for any \( q \in \mathbb{N} \), \( (X - x)^q \) in \( \Omega[X] \). Let \( g \in \Omega[X] \) be such that \( f = (X - x)^q g \). Apply \( \sigma' (\sigma: L \to \Omega) \) extends to a \( K \)-isomorphism \( \sigma_0 : \Omega \to \Omega \), and \( \sigma_0 \) extends to a unique \( K \)-algebra isomorphism \( \sigma' : \Omega[X] \to \Omega[X] \). Let \( \sigma' = \sigma(x) \). This implies that: for any \( q \in \mathbb{N} \), \( (X - x)^q | f \Leftrightarrow (X - \alpha x)^q | f \), which proves the claim. In conclusion, \( f = (\prod_{\sigma \in H} (X - \sigma x))^m \).

Comparing the degrees, we obtain \( [L : K] = mn = m[L : K] \). So, \( m = [L : K] = [K(x) : K] \), the inseparability degree of \( L \) over \( K \).

In the general case, each \( \sigma \in \text{Hom}_K(K(x), \Omega) \) has \( r := [L : K(x)] \), extensions to \( L \) (see V.3.11). So,
\[
\prod_{\sigma \in \text{Hom}_L(L, \Omega)} (X - \sigma x) = \prod_{\tau \in \text{Hom}_K(K(x), \Omega)} (X - \tau x)^r.
\]

By (1), we have:
\[
P(x, L/K) = f^{[L : K(x)]} = \left( \prod_{\tau \in \text{Hom}_K(K(x), \Omega)} (X - \tau x)^m \right)^{[L : K(x)]} = \left( \prod_{\sigma \in \text{Hom}_L(L, \Omega)} (X - \sigma x) \right)^{m[L : K(x)]/r},
\]

where
\[
m[L : K(x)]/r = [K(x) : K] \cdot [L : K(x)]/[L : K(x)] = [K(x) : K] \cdot [L : K(x)],
\]
\[
= [L : K].
\]

This proves (2). The second part of (3) and (4) follows from (2). \( \square \)

### 2.5 Corollary

Let \( K \subseteq L \) be a Galois extension and let \( G \) be its Galois group. Then, for any \( x \in L \),
\[
\text{Tr}_{L/K}(x) = \sum_{\sigma \in G} \sigma x
\]
\[
\text{N}_{L/K}(x) = \prod_{\sigma \in G} \sigma x.
\]
In particular, for any $\sigma \in G$, $N_{L/K}(\sigma x) = N_{L/K}(x)$ and $\text{Tr}_{L/K}(\sigma x) = \text{Tr}_{L/K}(x)$.

In the case of separable extensions, $d = 1$ in (2), (3), (4). In fact the inseparable extensions are exactly the extensions for which $\text{Tr} = 0$:

2.6 Proposition. Let $K \subseteq L$ be a finite extension. Then $\text{Tr}_{L/K}: L \rightarrow K$ is the 0 mapping if and only if $K \subseteq L$ is inseparable.

Proof. If $K \subseteq L$ is inseparable, then $\text{char } K = p > 0$ and the inseparability degree $d = [L : K]$ is a power of $p$ (3.27). Formula (3) implies $\text{Tr}_{L/K} = 0$.

Conversely, suppose $K \subseteq L$ is separable. Then $\text{Hom}_K(L, \Omega)$ has $n = [L : K]$ elements $\sigma_1, \ldots, \sigma_n$ and $\text{Tr}_{L/K}(x) = \sigma_1 x + \ldots + \sigma_n x$. If $\text{Tr}_{L/K} = 0$, then $\sigma_1 + \ldots + \sigma_n = 0$, so $\sigma_1, \ldots, \sigma_n$ are linearly dependent, contradicting Dedekind's lemma.

2.7 Example. Let $K$ be a field and let $d \in K$. If $g = X^2 - d$ is irreducible in $K[X]$ ($\Leftrightarrow d$ is not a square in $K$) and $e = \sqrt{d}$ is a root of $g$ in $\Omega$, the extension $K \subseteq K(e) =: L$ has degree 2 and it is normal. So, $\text{Hom}_K(L, \Omega) = \text{Hom}_K(L, L) = G(L/K)$. If $\text{char } K = 2$, then $g$ is inseparable and $G(L/K) = \{\text{id}, \sigma\}$, where $\sigma(e) = -e$. A basis is $\{1, e\}$. From (3) and (4) we have, for any $a, b \in K$:

- if $\text{char } K = 2$:
  $N_{L/K}(a + be) = (a + be)^2 = a^2 + b^2 d$, $\text{Tr}_{L/K}(a + be) = 2(a + be) = 0$;
- if $\text{char } K \neq 2$:
  $N_{L/K}(a + be) = (a + be)(a - be) = a^2 - b^2 d$, $\text{Tr}_{L/K}(a + be) = 2a$.

An important property of the trace and the norm is the transitivity property: If $R \subseteq S \subseteq T$ are extensions of commutative rings such that $S$
VI.2 Trace and norm

is a free $R$-module and $T$ is a free $S$-module (of finite ranks), then $T$ is a free $R$-module of finite rank\(^{19}\) and, for any $x \in T$:

$$N_{T/R}(x) = N_{S/R}(N_{T/S}(x)), \quad \text{Tr}_{T/R}(x) = \text{Tr}_{S/R}(\text{Tr}_{T/S}(x)).$$

We shall prove this for field extensions; a proof for the general case uses linear algebra methods.

2.8 Theorem. (The transitivity of the trace and norm) Let $K \subseteq L \subseteq E$ be finite extensions of fields. Then:

$$\text{Tr}_{E/K} = \text{Tr}_{L/K} \circ \text{Tr}_{E/L} \quad \text{and} \quad N_{E/K} = N_{L/K} \circ N_{E/L}.$$

Proof. We use 2.4.(3), 2.4.(4).

There exists a bijection

$$\Psi : \text{Hom}_K(L, \Omega) \times \text{Hom}_L(E, \Omega) \to \text{Hom}_K(E, \Omega),$$

defined as follows: for any $\sigma \in \text{Hom}_K(L, \Omega)$, fix a $K$-automorphism $\overline{\sigma} : \Omega \to \Omega$ with $\overline{\sigma}|_L = \sigma$. For any $(\sigma, \tau) \in \text{Hom}_K(L, \Omega) \times \text{Hom}_L(E, \Omega)$, define $\Psi(\sigma, \tau) := \overline{\sigma} \circ \tau$ (which is a $K$-homomorphism from $E$ to $\Omega$ since $\overline{\sigma}$ and $\tau$ are $K$-homomorphisms). We show that $\Psi$ is surjective. If $\eta : E \to \Omega$ is a $K$-homomorphism, let $\sigma := \eta|_L \in \text{Hom}_K(L, \Omega)$. Let $\tau := \overline{\sigma}^{-1} \circ \eta : E \to \Omega$; $\tau$ is an $L$-homomorphism : $\forall x \in L$, $\tau(x) = \overline{\sigma}^{-1}(\eta(x)) = \overline{\sigma}^{-1}(\sigma(x)) = x$. Moreover, $\eta = \Psi(\sigma, \tau)$, since, $\forall y \in E$, $\overline{\sigma}\tau(y) = \overline{\sigma}\overline{\sigma}^{-1}\eta(y) = \eta(y)$. Since $\text{Hom}_K(L, \Omega) \times \text{Hom}_L(E, \Omega)$ and $\text{Hom}_K(E, \Omega)$ are finite and have the same cardinal (see V.3.11), $\Psi$ is also injective.

Let $H := \text{Hom}_K(L, \Omega)$, $G := \text{Hom}_L(E, \Omega)$, $J := \text{Hom}_K(E, \Omega)$. We prove for the trace $\text{Tr}$, the proof for the norm being similar. We have, $\forall x \in E$:

$$\text{Tr}_{L/K}(\text{Tr}_{E/L}(x)) = [L : K];(\sum_{\sigma \in H} \sigma([E : L];(\sum_{\tau \in G} \tau x)))$$

$$= [E : L][L : K];(\sum_{(\sigma, \eta) \in H \times G} \sigma \eta x)$$

$$= [E : K];(\sum_{\eta \in J} \eta x) = \text{Tr}_{E/K}(x). \quad \square$$

\(^{19}\) This fact is proven exactly as the theorem of transitivity of finite extensions.
Exercises

1. Let $K \subseteq L$ be an extension of finite fields, $|K| = q$, $|L| = q^n$.
   a) Write explicit formulas for $N_{L/K}(x)$ and $\text{Tr}_{L/K}(x)$, $\forall x \in L$.
   b) Prove that $\text{Tr} : L \rightarrow K$ is surjective. Can the hypothesis “$L$ is a finite field” be weakened?
   c) Prove that the norm $N_{L/K}$ is a homomorphism of multiplicative groups $N : L^* \rightarrow K^*$.
   d) If $\alpha$ is a generator of the cyclic group $L^*$, then $N(\alpha)$ is a generator of $K^*$. Deduce that the norm $N : L \rightarrow K$ is surjective.
   e) Calculate the cardinal of the set $\{x \in L \mid N_{L/K}(x) = 1\}$.

2. Let $n > 1$ and let $\omega$ be a primitive complex $n$th root of unity. Show that $N_{L/\mathbb{Q}}(\omega) = 1$, where $L = \mathbb{Q}(\omega)$.

3. Write down formulas for the trace and the norm of an arbitrary element of a quadratic extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{d})$, with $d \in \mathbb{Z}$ squarefree. The same problem for $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2})$.

4. Let $K \subseteq L$ be a finite extension and let $x \in L$ such that $L = K(x)$ and $\text{Irr}(x, K) = X^n + a_{n-1}X^{n-1} + \ldots + a_0$.
   Then $N_{L/K}(x) = (-1)^n a_0$ and $\text{Tr}_{L/K}(x) = -a_{n-1}$.

5. Let $p > 2$ be a prime and let $\omega$ be a primitive complex $p$th root of unity. Show that $N_{L/\mathbb{Q}}(1 - \omega) = p$, where $L = \mathbb{Q}(\omega)$. Compute $N_{L/\mathbb{Q}}(a - \omega)$, where $a \in \mathbb{Q}$. (Hint. Find $\text{Irr}(1 - \omega, \mathbb{Q})$).

6. Let $L/K$ be finite Galois. If $K \subseteq F \subseteq L$ is an intermediate field, then $F = K(\{\text{Tr}_{L/F}(x) \mid x \in L\})$. Is it true in general that $F = K(\{N_{L/F}(x) \mid x \in L\})$?

7. Let $K$ be a field and let $K(X)$ be the field of fractions of $K[X]$. Let $\varphi : K(X) \rightarrow K(X)$ be the unique $K$-homomorphism with $\varphi(X) = 1/(1 - X)$. Prove that $\varphi \in \text{Aut}_K(K(X))$. Determine the fixed field of $\varphi$. (Ind. $\varphi^3 = \text{id}$. Use the trace to find an element in the fixed field.)
8. The same problem, for \( \varphi(X) = 1/X \), respectively \( \psi(X) = 1 - X \). Determine the subfield of \( K(X) \) fixed by \( \varphi \) and \( \psi \).

### VI.3 Cyclic extensions and Kummer extensions

We fix a field \( K \), an algebraic closure \( \Omega \) of \( K \) and we suppose that all the algebraic extensions of \( K \) are subfields of \( \Omega \). We study first the splitting field of the polynomial \( X^n - 1 \) over \( K \).

Let \( L \) be an extension of \( K \). Recall some notations and facts from IV.3: \( \mathbb{U}_n(L) := \{ x \in L \mid x^n = 1 \} \) is a finite subgroup (necessarily cyclic) of \( (L^*, \cdot) \); if \( |\mathbb{U}_n(L)| = n \), then any generator of \( \mathbb{U}_n(L) \) is called a primitive \( n \)th root of unity. A primitive \( n \)th root of unity exists in \( \Omega \) \( \iff \) \( |\mathbb{U}_n(L)| = n \) \( \iff \) char \( K \)-\text{div} \( n \).

\( \mathbb{P}_n(L) := \{ \omega \in \mathbb{U}_n(L) \mid \text{ord } \omega = n \} \) is the set of primitive \( n \)th roots of unity in \( L \). If \( L \) contains a primitive \( n \)th root of unity \( \omega \), then \( \mathbb{P}_n(L) = \{ \omega^i \mid 1 \leq i < n, \ (i, n) = 1 \} \), so \( |\mathbb{P}_n(L)| = \varphi(n) \). The splitting field of \( X^n - 1 \) over \( K \) is also called the \( n \)-th cyclotomic field over \( K \).

**3.1 Proposition.** Suppose \( n \in \mathbb{N}^* \), \( K \) is a field with char \( K \)-\text{div} \( n \) and \( L \) is a splitting field of \( X^n - 1 \) over \( K \). Let \( \omega \) be a primitive \( n \)th root of unity in \( L \). Then \( L = K(\omega) \), \( K \subseteq L \) is a Galois extension and its Galois group \( G \) is isomorphic to a subgroup of \( (\mathbb{U}(\mathbb{Z}/n\mathbb{Z}), \cdot) \), the group of invertible elements of \( \mathbb{Z}/n\mathbb{Z} \). In particular, \( G \) is Abelian and \( [L : K] \) is a divisor of \( \varphi(n) \).

**Proof.** \( K \subseteq L \) is a separable extension, since \( (X^n - 1)' = nX^{n-1} \neq 0 \) is relatively prime to \( X^n - 1 \). The extension \( K \subseteq L \) is also normal, being a splitting field. If \( \omega \) generates \( \mathbb{U}_n(L) \), then \( \mathbb{U}_n(L) = \{ \omega^i \mid 0 \leq i < n \} \) and clearly \( L = K(\omega) \). For any \( \sigma \in G, \sigma\omega \in \mathbb{U}_n(L) \) (\( \sigma \) permutes the roots of \( X^n - 1 \)). For any \( k \in \mathbb{N}, \omega^k = 1 \iff \sigma(\omega^k) = 1 \iff (\sigma\omega)^k = 1 \). Thus, the
order of $\sigma \omega$ in $U_n(L)$ is equal to $\text{ord} \omega = n$, whence $\sigma \omega \in P_n(L)$. But $P_n(L) = \{ \omega^i \mid 0 \leq i < n, (i, n) = 1 \}$. So, $\forall \sigma \in G, \exists! i < n, (i, n) = 1$, such that $\sigma \omega = \omega^i$. Define then $\psi(\sigma) = i + n\mathbb{Z}$. We obtain a function $\psi : G \to U(\mathbb{Z}/n\mathbb{Z})$. An easy check shows that $\psi$ is a homomorphism; $\psi$ is injective, since $\psi(\sigma) = 1 + n\mathbb{Z}$ implies $\sigma \omega = \omega \Rightarrow \sigma = \text{id}$. Since $U(\mathbb{Z}/n\mathbb{Z}) = \{ i + n\mathbb{Z} \mid 0 \leq i < n, (i, n) = 1 \}$ is Abelian and has $\varphi(n)$ elements, the other claims are evident.

3.2 Remark. a) If $K$ already contains some $n$th roots of unity, and $\omega$ is a primitive $n$th root of unity, then we can have $[K(\omega) : K] < \varphi(n)$. For instance, if $n = 8$ and $K = \mathbb{Q}(i)$ ($i^2 = -1$), then $(\omega^2)^4 = 1$, so $\omega^2 = \pm i$. Thus $\text{Irr}(\omega, K) = X^2 - i$. So, $[K(\omega) : K] = 2 < \varphi(8) = 4$.

b) If $K = \mathbb{Q}$, then $G(\mathbb{Q}(\omega)/\mathbb{Q}) \cong U(\mathbb{Z}/n\mathbb{Z})$. Indeed, $\text{Irr}(\omega, K) = \Phi_n$ (the $n$th cyclotomic polynomial) whose degree is $\varphi(n)$. So, $|G(\mathbb{Q}(\omega)/\mathbb{Q})| = [\mathbb{Q}(\omega) : \mathbb{Q}] = \varphi(n)$ and $\psi$ in the proof must be an isomorphism.

The extensions of the form $K \subseteq K(a)$, where $a$ has the property that $a^n = b \in K$ for some $n \in \mathbb{N}$, are very important in the problem of solvability by radicals of polynomial equations. A suggestive writing for this situation is $a = \sqrt[n]{b}$ or $a = b^{1/n}$ and an extension as above is written $K \subseteq K(\sqrt[n]{b})$. These extensions have a simple description if $K$ contains a primitive $n$th root of unity.

3.3 Proposition. Let $K$ be a field containing a primitive $n$th root of unity $\omega$. Let $b \in K^*$ and let $L = K(a)$, where $a$ is a root of $X^n - b$ (in $\Omega$). Then:

a) $K \subseteq L$ is a Galois extension and $G(L/K)$ is cyclic of order $m$, with $m \mid n$.

b) $[L : K] = m$ is the order of $bK^*/n$ in the group $(K^*/K^*/n, \cdot)$. $(K^*/n := \{ x^n \mid x \in K^* \}$ is a subgroup in $(K^*, \cdot))$

c) $\text{Irr}(a, K) = X^m - c$, for some $c \in K$. 
VI.3 Cyclic extensions and Kummer extensions

Proof. a) The roots of \(X^n - b\) are \(\{a\omega^i \mid 0 \leq i < n\}\), thus \(K \subseteq L\) is normal. It is also separable (because \(\text{char } K\) does not divide \(n\)), so it is Galois. If \(\sigma \in G := G(L/K)\), then \(\sigma(a) = a\omega^i\); define \(\varphi : G \to \mathbb{Z}/n\mathbb{Z}\), \(\varphi(\sigma) = i + n\mathbb{Z}\). The function \(\varphi\) is a injective homomorphism, since \(\varphi(\sigma) = 0 + n\mathbb{Z}\) implies \(\sigma a = a \Rightarrow \sigma = \text{id}\). Thus, \(G\) is isomorphic to \(\text{Im } \varphi\), a subgroup of \(\mathbb{Z}/n\mathbb{Z}\), so it is cyclic. \(|G|\) divides \(n\), by the theorem of Lagrange.

b) Let \(\eta\) be a generator of \(G\). Then \(m = |G| = \text{ord } \eta\). There exists \(s < n\) such that \(\eta(a) = a\omega^s\). We have
\[
N_{L/K}(a) = \prod_{\sigma \in G} \sigma a = \prod_{0 \leq i < m} a\omega^i = a^m\omega^i \in K,
\]
with \(t \in \mathbb{N}\), so \(a^m \in K\). Therefore, \(b^m = (a^n)^m = (a^m)^n \in K^*\). Let \(q\) denote the order of \(bK^*/nK^*\) in grupul \((K^*/nK^*, \cdot)\). We have \(q \mid m\). Let us show that \(m \mid q\). Since \(b^q \in K^*\), \(b^q = c^n\), for some \(c \in K\). So, \((a^q)^n = b^q = c^n\), whence \(a^q = c\omega^i\), with \(i < n\). Thus, \(a^q \in K\), so \(a^q\) is fixed by \(\eta\). \(\eta(a^q) = \eta(a)^q = a^q\omega^q = a^q\). We obtain \(\omega^sq = 1 \Rightarrow \eta^q = \text{id}\). This shows that \(\text{ord } \eta = m\) divides \(q\). From \(m \mid q\) and \(q \mid m\) we deduce that \(m = q\).

c) \(\deg \text{Irr}(a, K) = [L : K] = m\), and \(a\) is a root of \(X^m - a^m \in K[X]\). \(\square\)

3.4 Definition. A Galois extension (normal and separable) whose Galois group is cyclic (respectively Abelian) is called a cyclic extension (respectively Abelian extension).

The previous proposition says: if \(K\) contains a primitive nth root of unity and \(b \in K^*\), then \(K \subseteq K(\sqrt[n]{b})\) is cyclic.

We want to prove that, conversely, any finite cyclic extension of \(K\) is of the form \(K \subseteq K(\sqrt[n]{b})\) (if \(K\) contains a primitive nth root of unity). This step is crucial in the proof of the characterization of solvability by radicals of an equation. We use the following result (very important by itself):
3.5 Lemma. (Hilbert Satz 90)\(^{20}\) Let \( K \subseteq L \) be a cyclic extension of degree \( n \), let \( \sigma \) be a generator of \( \text{Gal}(L/K) \) and let \( x \in L^* \). Then:
\[
N_{L/K}(x) = 1 \iff \text{there exists } y \in L^* \text{ such that } x = \sigma(y)/y.
\]

Proof. One implication is easy: if \( x = \sigma(y)/y \), using \( N_{L/K}(y) = N_{L/K}(\sigma y) \) we obtain that \( N_{L/K}(x) = N_{L/K}(\sigma y)/N_{L/K}(y) = 1 \).

Suppose now that \( N_{L/K}(x) = x \sigma(x) \ldots \sigma^{n-1}(x) = 1 \). Let \( \eta := \sigma^{-1} \) and let \( x_0, \ldots, x_{n-1} \in L \), defined as follows: \( x_0 := x, \ x_1 := x \eta(x_0) = x \eta(x), \ldots, \ x_{n-1} := x \eta(x_{n-2}) = x \eta(x) \ldots \eta^{n-1}(x) \).

Let \( u := x_0 \cdot \text{id} + x_1 \cdot \eta + \ldots + x_{n-1} \eta^{n-1} \) ( "the Lagrange-Hilbert resolvent").

Consider \( u : L \to L \) as an endomorphism of the \( K \)-vector space \( L \). The homomorphisms \( \text{id}, \eta, \ldots, \eta^{n-1} \) from \( L^* \) to \( L^* \) are distinct. Dede-kind's lemma says they are linearly independent in the \( L \)-vector space \( L^* \); since \( x_0 \neq 0 \), we have \( u \neq 0 \), so there exists \( z \in L \) such that \( u(z) =: t \neq 0 \). Thus, \( t = u(z) = x_0 z + x_1 \eta(z) + \ldots + x_{n-1} \eta^{n-1}(z) \), so
\[
\eta(t) = \eta(x_0) \eta(z) + \eta(x_1) \eta^2(z) + \ldots + \eta(x_{n-1}) \eta^n(z) = x_1 \eta(z) + x_2 \eta^2(z) + \ldots + z \text{ (we used that } x_{n-1} = N_{L/K}(x) = 1 \text{ and } \eta^n = \text{id})
\]
We remark that \( \eta(t) = t/x_0 \), so \( x = x_0 = t/\eta(t) \). Setting \( y := \sigma(t) \) finishes the proof.

\[\square\]

3.6 Proposition. Let \( K \) be a field containing a primitive \( n \)th root of unity \( \omega \) and let \( K \subseteq L \) be a cyclic extension of degree \( n \). Then there exists \( a \in L^* \), with \( a^n \in K \), such that \( L = K(a) \).

Proof. Let \( G := \text{Gal}(L/K) \) and let \( \sigma \) be a generator of \( G \). Since \( \omega \in K \), \( N_{L/K}(\omega) = \omega^n = 1 \). By Hilbert's Satz 90, there exists \( a \in L \) with \( \sigma(a) = \omega a \), so \( \sigma^i(a) = \omega^i a \), \( 0 \leq i \leq n - 1 \). Thus, \( \{ \omega^i a \mid 0 \leq i \leq n - 1 \} \) are \( n \) distinct conjugates of \( a \) over \( K \), so \( \text{Irr}(a, K) \) has degree \( \geq n \). Because

\[\phantom{=}\]

\(^{20}\)This result is included as "Satz 90" in the Hilbert's monumental Bericht über die Theorie der algebraischen Zahlkörper (Zahlbericht for short, published in 1897).
\[ \sigma(a^n) = \omega^n a^n = a^n, \] a \text{ is fixed by the subgroup } \langle \sigma \rangle = G, \text{ so } a^n \in K. \] The polynomial \( X^n - a^n \in K[X] \) has the root \( a \) and its degree is \( n \), so it is equal to \( \text{Irr}(a, K) \). Thus, \( [K(a) : K] = n \) and \( K(a) = L. \)

The propositions 3.3 and 3.6 determine all cyclic extensions of degree \( n \) of a field \( K \) of characteristic exponent \( p \), such that \( p \nmid n \) and a \( K \) contains a primitive \( n \)th root of unity. The cyclic extensions of degree \( p \) of a field of characteristic \( p > 0 \) are described in what follows. This part is not used in the theory of solvability by radicals and may be skipped in a first reading.

### 3.7 Lemma. (Hilbert Satz 90, additive version)
Let \( K \subseteq L \) be a cyclic extension of degree \( n \), let \( \sigma \) be a generator of \( G(L/K) \) and let \( x \in L^* \). Then:

\[ \text{Tr}_{L/K}(x) = 0 \iff \text{there exists } y \in L^* \text{ such that } x = \sigma(y) - y. \]

**Proof.** The argument is essentially the same as in the multiplicative version (3.5). If \( x = \sigma(y) - y \), then \( \text{Tr}_{L/K}(x) = \text{Tr}_{L/K}(\sigma y) - \text{Tr}_{L/K}(y) = 0 \), since \( \text{Tr}_{L/K}(y) = \text{Tr}_{L/K}(\sigma y) \).

Suppose \( \text{Tr}_{L/K}(x) = x + \sigma(x) + \ldots + \sigma^{n-1}(x) = 0 \). Let \( x_0, \ldots, x_{n-1} \in L \), defined as follows:

\[ x_0 := x, \quad x_1 := x + \sigma(x_0) = x + \sigma(x), \ldots, \]
\[ x_{n-1} := x + \sigma(x_{n-2}) = x + \sigma(x) + \ldots + \sigma^{n-1}(x) (= \text{Tr}_{L/K}(x) = 0) \]

Let \( u = x_0 \sigma + x_1 \sigma^2 + \ldots + x_{n-2} \sigma^{n-1} \in \text{End}_K(L) \). We have:

\[ \sigma u = \sigma(x_0) \cdot \sigma^2 + \sigma(x_1) \cdot \sigma^3 + \ldots + \sigma(x_{n-2}) \cdot \sigma^n \]
\[ = (x_1 - x) \cdot \sigma^2 + (x_2 - x) \cdot \sigma^3 + \ldots + (x_{n-1} - x) \cdot \sigma^n \]
\[ = x_1 \cdot \sigma^2 + x_2 \cdot \sigma^3 + \ldots + x_{n-1} \cdot \sigma^n + x \cdot \sigma - x \cdot (\sigma + \sigma^2 + \ldots + \sigma^{n-1} + \sigma^n) \]
\[ = u - x \cdot \text{Tr}_{L/K}. \]

We have \( \text{Tr}_{L/K} = \text{id} + \sigma + \ldots + \sigma^{n-1} \neq 0 \) since, by Dedekind's Lemma, \( \text{id} \), \( \sigma \), \( \ldots \), \( \sigma^{n-1} \) are \( L \)-linearly independent. Let \( z \in L^* \) such that \( \text{Tr}_{L/K}(z) \neq 0 \). If \( t := z / \text{Tr}_{L/K}(z) \), then \( \text{Tr}_{L/K}(t) = 1 \) and thus \( \sigma(u(t)) = u(t) - x \). Thus, \( x = \sigma(y) - y \), where \( y = -u(t) \). \qed
3.8 Theorem. (Artin-Schreier\textsuperscript{21}) Let $K$ be a field of characteristic $p > 0$ and let $K \subseteq L$ be a cyclic extension of degree $p$. Then there exists $a \in L$, with $a^p - a = b \in K$, such that $L = K(a)$.

Conversely, let $b \in K$ such that $f = X^p - X - b$ does not split over $K$. Then $f$ is irreducible in $K[X]$ and, for any root $a \in \Omega$ of $f$, the extension $K \subseteq K(a)$ is cyclic of degree $p$.

Proof. Note that it is enough to suppose that $K \subseteq L$ is Galois; its Galois group must then be cyclic, having prime order.

Let $\sigma$ be a generator of $G := \text{G}(L/K)$. We have $\text{Tr}_{L/K}(1) = [L : K] \cdot 1 = p \cdot 1 = 0$. Applying the additive Hilbert Satz 90, there exists $a \in L$ such that $1 = \sigma a - a$, so $\sigma a = a + 1$. Then $\sigma^2 a = a + 2$ and, by induction, $\sigma^n a = a + n \cdot 1$, for any $n \in \mathbb{N}$. Thus, $a$ has $p$ distinct conjugates in $L$, namely $a$, $\sigma a = a + 1$, ..., $\sigma^{p-1} a = a + (p - 1)$, so $\text{Irr}(a, K)$ has degree $> p$. Since $K(a) \subseteq L$ and $[L : K] = p$, we have $K(a) = L$. Because $\sigma(a^p - a) = (\sigma a)^p - \sigma a = (a + 1)^p - a - 1 = a^p - a$, $a^p - a$ is fixed by $\sigma$, whence $a^p - a \in L^{<\sigma>} = K$.

Conversely, let $b \in K$, $f = X^p - X - b$ and let $a \in \Omega \setminus K$ be a root of $f$. Then $a + 1$ is also a root: $(a + 1)^p - (a + 1) - b = a^p - a - b = 0$. Then there exists a $K$-isomorphism $\sigma : K(a) \to K(a + 1)$, with $\sigma(a) = a + 1$. Therefore $\sigma \in \text{G}(K(a)/K)$ and $\text{ord} \sigma = p$, since $\sigma^n a = a + n \cdot 1$, for any $n \in \mathbb{N}$. Because $|\text{G}(K(a)/K)| = p \geq [K(a) : K]$, $K \subseteq K(a)$ is Galois and $p = [K(a) : K]$ (use V.4.4). Since $\text{deg Irr}(a, K) = p = \text{deg} f$, we deduce that $f = \text{Irr}(a, K)$, so $f$ is irreducible. $\text{G}(K(a)/K)$ has order $p$, so it is cyclic.

3.9 Remark. The proof also shows that, for any $b \in K$, where char $p > 0$, $X^p - X - b$ is either irreducible in $K[X]$ or splits over $K$.

\textsuperscript{21} Otto Schreier (1901-1929), German mathematician.
The theorems that characterize cyclic extensions, combined with the fact that any Abelian finite group is a product of cyclic groups, lead to *structure theorems of finite Abelian extensions of the field K*. The theory we developed so far allows us to obtain such theorems in two situations (described by the hypotheses in prop. 3.6, respectively 3.8):

- for Abelian extensions of exponent \( n \), if \( K \) contains a primitive \( n \)th root of unity;
- for Abelian extensions of exponent \( p \), if \( \text{char } K = p \).

This type of theorems are known as *Kummer Theory*.\(^{22}\)

If \( K \) does not contain a primitive \( n \)th root of unity, the theory is considerably more complicated. In some cases (when \( K \) is a “local field” or when \( K \) is a finite extension of \( \mathbb{Q} \)), a theory that describes all Abelian extensions of \( K \) using only the internal structure of \( K \) is the *Class Field Theory*, one of the major accomplishments of 20th century Algebra. (see e.g. *Neukirch [1986]*). We describe now *Kummer theory*.

3.10 **Definition.** If \( G \) is a group such that \( \{ \text{ord} x \mid x \in G \} \) is a finite set, then the LCM of \( \{ \text{ord} x \mid x \in G \} \) is called the *exponent* of \( G \), denoted \( \text{exp } G \). If \( G \) is finite, then its exponent is defined and \( \text{exp } G \) divides \( |G| \), by Lagrange's theorem. If \( K \subseteq L \) is a finite Galois extension, the *exponent* of the extension is the exponent of its Galois group.

Let \( n \in \mathbb{N}^* \) and let \( K \) be a field containing a primitive \( n \)th root of unity. An *\( n \)-Kummer extension* is a finite Galois extension \( K \subseteq L \), such that \( G(L/K) \) is an Abelian group whose exponent divides \( n \). An extension is called a *Kummer extension* if it is \( n \)-Kummer for some \( n \in \mathbb{N}^* \).

---

\(^{22}\) Ernst Eduard Kummer (1810-1893), German mathematician.
If \( K \) contains a primitive \( n \)th root of unity and \( b \in K^* \), 3.6 says that \( K(\sqrt[n]{b})/K \) is a cyclic extension and its degree divides \( n \), so it is an \( n \)-Kummer extension.

3.11 Lemma. Let \( L/K \) be an Abelian finite extension of exponent \( n \). Then \( L \) is the composite of a finite set of cyclic intermediate fields \( K \subseteq L_i, 1 \leq i \leq m, \) and each degree \([L_i : K]\) divides \( n \).

Proof. Let \( G := G(L/K) \). The finite Abelian group \( G \) is a direct product of cyclic subgroups: \( G = C_1 \times \ldots \times C_m \); since \( \exp G = n = |C_1| \ldots |C_m| \), it follows that \( |C_i| \) divides \( n \), \( 1 \leq i \leq m \). Take the subgroup \( H_i := \prod_{j \neq i} C_j \) and let \( L_i \) be the fixed subfield of \( H_i \). The extension \( L_i/K \) is Galois and \( G(L_i/K) \cong G/H_i \cong C_i \), so \( L_i/K \) is cyclic and \([L_i : K]\) divides \( n \). Let \( E \) be the composite \( L_1 \ldots L_m \). We have \( G(L/E) = H_1 \cap \ldots \cap H_m = \{ \text{id} \} \), so \( L = E \). \( \square \)

For any extension \( K \subseteq L \) and any subset \( S \) of \( K \), let \( \sqrt[n]{S} := \{ x \in L \mid x^n \in S \} \).

3.12 Proposition. Let \( n \in \mathbb{N}^* \) and let \( K \) be a field containing a primitive \( n \)th root of unity \( \omega \). Then a finite extension \( K \subseteq L \) is \( n \)-Kummer if and only if there exist \( a_1, \ldots, a_m \in K^* \setminus K^*n \) such that \( L = K(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m}) \). For such an extension, \( L = K(\sqrt[n]{S}) \), where \( S := L^n \cap K \) is a subgroup in \((K^*, \cdot)\).

Proof. a) Let \( L = K(b_1, \ldots, b_m) \), with \( b_i \in L \) and \( b_i^n = a_i, 1 \leq i \leq m \). Since \( \omega \in K \), \( X^n - a_i \) has \( n \) distinct roots in \( L \): \( b_i, b_i\omega, \ldots, b_i\omega^{n-1} \). So, \( L \) is a splitting field over \( K \) of the separable polynomial \( \prod_{1 \leq i \leq m}(X^n - a_i) \), which implies that \( K \subseteq L \) is Galois. If \( \sigma \in G(L/K) \), \( \sigma \) is determined by its values on the generators \( \{b_1, \ldots, b_m\} \). But \( \sigma(b_i) = b_i\omega^j \), with \( j \in \mathbb{N} \) depending on \( \sigma \) and \( i \). Let \( \tau \in G(L/K) \), with \( \tau(b_i) = b_i\omega^k \). Then:

\[
(\sigma \circ \tau)(b_i) = \sigma(b_i\omega^k) = \sigma(b_i)\sigma(\omega^k) = b_i\omega^j\omega^k = b_i\omega^{j+k} = (\tau \circ \sigma)(b_i), \forall i,
\]
so \( \sigma \circ \tau = \tau \circ \sigma \). Thus, \( G(L/K) \) is Abelian. Also, \( \sigma^n(b_i) = b_i \omega^{jn} = b_i \), so \( \sigma^n = \text{id} \) and the exponent of \( G(L/K) \) divides \( n \).

Suppose now that \( K \subseteq L \) is \( n \)-Kummer. By 3.11, \( L = L_1 \ldots L_m \), where each \( L_i \) is a cyclic extension of \( K \), included in \( L \); \( G(L_i/K) \) has order \( d_i \), with \( d_i | n \). For the extension \( K \subseteq L_i \) we can apply prop. 3.6, since \( \omega^{d_i} \) is a primitive \( d_i \)-th root of unity in \( K \). So, \( L_i = K(b_i) \), with \( b_i \in L^* \) and \( b_i^{d_i} = a_i \in K^* \). Then \( L = L(b_1, \ldots, b_m) \) and clearly \( b_i^n = a_i \in K^* \). We have \( d_i > 1 \) (otherwise \( K = L_i \)), so \( a_i \notin K^{*n} \) (if \( a_i = c^n \), for some \( c \in K \), then, for some \( j \), \( b_i = c \omega^j \in K \) and \( L_i = K(b_i) = K \), absurd).

The last statement is justified as follows: \( S \) is a subgroup in \( (K^*, \cdot) \) and \( K(\sqrt[n]{S}) \subseteq L \). But \( L = K(b_1, \ldots, b_m) \), with \( b_i \in L^* \) and \( b_i^n = a_i \in K^* \), \( 1 \leq i \leq m \), so \( a_i \in S \). So, \( L \subseteq K(\sqrt[n]{S}) \).

This result generalizes 3.3 and 3.6. In order to investigate the structure of the Galois group of an \( n \)-Kummer extension, we need some background.

3.13 Definition. Let \( (H, \cdot) \) be an Abelian group with the operation denoted multiplicatively, let \( 1 \) be its identity element and let \( n \in \mathbb{N}^* \). Let \( t_n(H) := \{ x \in H \mid x^n = 1 \} \).

An easy check shows that \( t_n(H) \) is a subgroup in \( H \) (it is the kernel of the homomorphism \( x \mapsto x^n \)) and \( \exp t_n(H) \) divides \( n \). We use this notion for an extension \( K \subseteq L \), for the factor group \( L^*/K^* \) (\( L^* \) is a group with respect to multiplication; \( K^* \) is a subgroup). Then:
\[
t_n(L^*/K^*) = \{ xK^* \mid x \in L^*, x^n \in K^* \}.
\]

3.14 Proposition. Let \( K \subseteq L \) be an \( n \)-Kummer extension. Then the canonical group homomorphism
\[
\phi : t_n(L^*/K^*) \to K^*/K^{*n}, \phi(xK^*) = x^n K^{*n}, \forall x \in L^* \text{ with } x^n \in K^*,
\]
is injective. The homomorphism \( \phi \) induces an isomorphism \( t_n(L^*/K^*) \cong \text{Im} \phi = (L^{*n} \cap K^*)/K^{*n} \), which is a subgroup in \( K^*/K^{*n} \).
VI. Applications of Galois Theory

Proof. \( \varphi \) is well defined: if \( xK^n \in t_n(L^*/K^*) \), then \( x^n \in K^* \); if \( x, y \in L^* \) such that \( xK^n = yK^n \in t_n(L^*/K^*) \), then \( xy^{-1} \in K^* \), so \( (xy^{-1})^n \in K^{*n} \), i.e. \( x^nK^n = y^nK^n \). It is immediate that \( \varphi \) preserves multiplication.

Let us show \( \text{Ker} \varphi = \{1K^*\} \). Let \( x^nK^n = 1K^n \) for \( x \in L^* \) with \( x^n \in K \). Hence \( x^n \in K^{*n} \), so there exists \( a \in K^* \) such that \( x^n = a^n \). Then \( x = a\zeta \), where \( \zeta \) is an \( n \)th root of unity (\( \zeta \in K \) by hypothesis).

So, \( x \in K^* \) and \( xK^n = 1K^n \).

\( \text{Im} \varphi = \{x^nK^n | x \in L^*, x^n \in K^*\} = \{x^nK^n | x^n \in L^* \cap K^*\} = (L^* \cap K^*)/K^n. \)

The inverse of \( \varphi \) is

\[ \psi : (L^* \cap K^*)/K^n \rightarrow t_n(L^*/K^*), \psi(yK^n) = \sqrt[n]{y}K^n. \]

The definition of \( \psi \) is correct: \( \forall y \in L^* \cap K^*, \exists x := \sqrt[n]{y} \in L^* \) such that \( x^n = y \). The class \( xK^* \) is independent on the choice of the root \( x \) of \( X^n - y \), since any other root is of the form \( xa \), with \( a \in U_n(K) \), so \( xaK^* = xK^* \).

The main theorem says that the \( n \)-Kummer extensions of \( K \) are in one-to-one correspondence with the finite subgroups of \( K^*/K^{*n} \). If \( K \subseteq L \) is \( n \)-Kummer, then

\[ G(L/K) \cong t_n(L/K) \cong (L^* \cap K^*)/K^n. \]

We need some group theoretical facts, stated in the following lemma.

3.15 Lemma. Let \( (G, \cdot) \), \( (H, \cdot) \) be finite Abelian groups and let \( (C, \cdot) \) be a cyclic group. Denote by 1 the neutral elements.

a) \( \text{Hom}(G, C^*) \cong G \). (\( C^* \) is the multiplicative group of nonzero complex numbers). If \( \exp G \) divides \( |C| \), then

\[ \text{Hom}(G, C^*) \cong \text{Hom}(G, C) \cong G \].

\[ \text{Hom}(G, C^*) \) is called the dual of the group \( G \).
VI.3 Cyclic extensions and Kummer extensions

b) Let \( p : G \times H \to C \) be a bilinear function: for any \( a, b \in G, \forall x, y \in H \), we have:

\[
p(ab, x) = p(a, x)p(b, x), \quad (*)
\]

\[
p(a, xy) = p(a, x)p(a, y). \quad (**)\]

Then, for any \( x \in G \), \( p_x : G \to C \), \( p_x(a) = p(a, x), \forall a \in G \), is a group homomorphism. The function \( \eta : H \to \text{Hom}(G, C) \), \( \eta(x) = p_x, \forall x \in H \) is a group homomorphism.

c) Suppose that the bilinear function \( p \) is nondegenerate:

\[
\{x \in H \mid p(a, x) = 1, \forall a \in G\} = \{1\} \quad \text{and} \quad \{a \in G \mid p(a, x) = 1, \forall x \in H\} = \{1\}.
\]

Then \( \exp G \) divides \( |C| \) and \( \eta : H \to \text{Hom}(G, C) \) is an isomorphism, so:

\[
H \cong \text{Hom}(G, C) \cong G.
\]

**Proof.** Note first that, for any Abelian group \((A;\cdot)\), \( \text{Hom}(G, A) = \{\varphi : G \to A \mid \varphi \text{ is a homomorphism}\} \) is an Abelian group with respect to the law \((\alpha\beta)(x) = \alpha(x)\beta(x), \forall \alpha, \beta \in \text{Hom}(G, A), \forall x \in G\).

a) Let \( G' := \text{Hom}(G, \mathbb{C}^*) \). Let \( \exp G = n \). Then \( \alpha(x)^n = \alpha(x^n) = 1, \forall \alpha \in G', \forall x \in G \). So, \( \text{Im}\alpha \subseteq U_n \), the group of complex \( n \)th roots of unity, cyclic of order \( n \). Thus, \( G' = \text{Hom}(G, U_n) \). If \( n \) divides \( |C| \), then \( C \) includes a unique subgroup with \( n \) elements \( C_n \) (\( C_n \) is cyclic) and, as above, \( \text{Hom}(G, C) = \text{Hom}(G, C_n) \).

Since \( C_n \cong U_n \), we have \( \text{Hom}(G, C_n) \cong \text{Hom}(G, U_n) = G' \).

Let us prove that \( G' \cong G \). Suppose first that \( G \) is cyclic of order \( n \) and let \( g \in G \) be a generator. Let \( \alpha \in G' \) and let \( \omega \) generate \( U_n \). Then \( \alpha \) is determined by \( \alpha(g) \), and \( \alpha(g) = \omega^s \), for some \( s \in \mathbb{N} \). Define \( \varphi : \mathbb{Z} \to G', \varphi(s) = \alpha_s, \forall s \in \mathbb{Z} \), where \( \alpha_s(g) = \omega^s \). A quick check shows that \( \varphi \) is a surjective group homomorphism and \( \text{Ker}\varphi = n\mathbb{Z} \). So, \( \mathbb{Z}/n\mathbb{Z} \cong G' \), which means that \( G' \) is cyclic of order \( n \), isomorphic to \( G \).

If \( G \) is an arbitrary Abelian finite group, the invariant factors theorem says that \( G = G_1 \times \ldots \times G_m \), with \( G_i \) cyclic groups. But

\[
\text{Hom}(G_1 \times \ldots \times G_m, \mathbb{C}) \cong \text{Hom}(G_1, \mathbb{C}) \times \ldots \times \text{Hom}(G_m, \mathbb{C})
\]
VI. Applications of Galois Theory

(see II.3.20) and each \( \text{Hom}(G_i, \mathbb{C}) \cong G_i \), as shown. Thus \( \text{Hom}(G, \mathbb{C}) \cong G_1 \times \ldots \times G_m = G \).

b) Condition (*) means that \( p_a \) is a homomorphism. Let \( x, y \in H \) and let \( a \in G \). We have

\[
\eta(xy)(a) = p_{xy}(a) = p(a, xy) = p(a, x)p(a, y) = \eta(x)(a) \cdot \eta(y)(a),
\]

so \( \eta(xy) = \eta(x) \eta(y) \).

c) Ker \( \eta = \{ x \in H \mid \eta(x) = 1 \} = \{ x \in H \mid p(a, x) = 1, \forall a \in G \} = \{1 \} \) (\( p \) is nondegenerate). So \( \eta \) is injective. Let \( m = |C| \). For any \( x \in G \), \( p(x^m, y) = p(x, y)^m = 1, \forall y \in H \). Because \( p \) is nondegenerate, \( x^m = 1, \forall x \in G \). But exp \( G \mid m \). So, \( G \cong \text{Hom}(G, C) \) by a).

Therefore, \( H \cong \text{Im} \eta \leq \text{Hom}(G, C) \cong G \), so \( H \) is isomorphic to a subgroup of \( G \), in particular \( |H| \leq |G| \). The situation is symmetric in \( G \) and \( H \), so \( G \) is isomorphic to a subgroup of \( H \) and \( |G| \leq |H| \). But \( G \) and \( H \) are finite, so \( |G| = |H| \) and \( G \cong H \).

3.16 Theorem. (multiplicative Kummer Theory) Let \( K \) be a field containing a primitive \( n \)th root of unity \( \omega \).

a) If \( K \subseteq L \) is an \( n \)-Kummer extension, then there is a canonical isomorphism:

\[
t_n(L^*/K^*) \cong \text{Hom}(G(L/K), U_n)
\]

\[
aK^* \mapsto \chi_a \text{ where } \chi_a(\sigma) = \frac{\sigma(a)}{a},
\]

for any \( aK^* \in t_n(L^*/K^*) \), \( \forall \sigma \in G(L/K) \). Also, there is a canonical isomorphism

\[
(L^*/n \cap K^*)/K^* n \cong \text{Hom}(G(L/K), U_n),
\]

\[
bK^* n \mapsto \chi_b \text{ where } \chi_b(\sigma) = \sigma(\sqrt[n]{b})/\sqrt[n]{b},
\]

\( \forall bK^* n \in (L^*/n \cap K^*)/K^* n, \forall \sigma \in G(L/K) \).

Therefore, \( G(L/K) \) is isomorphic to the finite subgroup \( (L^*/n \cap K^*)/K^* n \) of \( K^*/K^* n \).

b) There exists an order preserving bijection between the set of all \( n \)-Kummer extensions of \( K \) included in \( \Omega \) and the finite subgroups of \( K^*/K^* n \):
VI.3 Cyclic extensions and Kummer extensions

the extension $K \subseteq L$ corresponds to the subgroup $(L^* \cap K^*)/K^*$; the subgroup $S/K^*$, with $K^* \subseteq S \subseteq K^*$, corresponds to the extension $K \subseteq K^{\sqrt[n]{S}}$.

The degree of the extension is equal to the order of its corresponding subgroup.

**Proof.** (a) Let $U_n = U_n(K)$ be the group of the $n$th roots of unity in $K$ (we know that $U_n = U_n(\Omega)$ is a cyclic group of order $n$). Define:

$$p : G(L/K) \times t_n(L^*/K^*) \to U_n, \quad p(\sigma, xK^*) := \frac{\sigma(x)}{x},$$

for any $\sigma \in G(L/K)$ and any $x \in L^*$ with $xK^* \in t_n(L^*/K^*)$.

The definition is correct: $\forall x \in L^*$ with $xK^* \in t_n(L^*/K^*)$, we have $x^n = a \in K$, so $\sigma(x)$ is a root of $X^n - a$, so $(\sigma(x)/x)^n = 1$; thus, $\sigma(x)/x \in U_n$. It is easily seen that $p(\sigma, xK^*)$ is independent on the choice of the representative of the class $xK^*$.

$p$ is bilinear. Indeed, $\forall \sigma, \tau \in G(L/K)$ and $\forall xK^*, yK^* \in t_n(L^*/K^*)$, we have:

$$p(\sigma \tau, xK^*) = \frac{\sigma \tau(x)}{x} = \frac{\sigma(x)}{x} \cdot \frac{\tau(x)}{x} = \frac{\sigma(x)}{x} \cdot \frac{\tau(x)}{x} = p(\sigma, xK^*)p(\tau, xK^*)$$

We used first that $\sigma \tau = \tau \sigma$ $(G(L/K)$ is Abelian), then that $\sigma(x)/x \in K$, so it is fixed by $\tau$. Also, $p(\sigma, xK^* \cdot yK^*) = p(\sigma, xyK^*) = \sigma(xy)/xy = (\sigma(x)/x) \cdot (\sigma(y)/y) = p(\sigma, xK^*)p(\sigma, yK^*)$.

$p$ is nondegenerate: if $\sigma \in G(L/K)$ and $p(\sigma, xK^*) = 1, \forall xK^* \in t_n(L^*/K^*)$, then $\sigma(x) = x, \forall x \in L$ with $x^n \in K$. But $L = K\{x \in L \mid x^n \in K\}$ (see 3.12), so $\sigma = \text{id}$. If $xK^* \in t_n(L^*/K^*)$ with $p(\sigma, xK^*) = 1, \forall \sigma \in G(L/K)$, then $x$ is fixed by any $\sigma \in G(L/K)$, so $x \in K$. Thus, $xK^* = 1K^*$.

We have $t_n(L^*/K^*) \cong \text{Hom}(G(L/K), U_n)$, via the isomorphism in Lemma 3.15.c). Using $\psi : (L^* \cap K^*)/K^* \to t_n(L^*/K^*)$, the isomorphism in 3.14, $\psi(yK^*) = \sqrt[n]{y}K^*$, we obtain the other isomorphism. Finally, 3.15.a) implies
\[ t_n(L^*/K^*) \cong (L^*/n \cap K^*)/K^* \cong G(L/K). \]

b) Any finite subgroup of \( K^*/K^* \) can be written uniquely as \( S/K^* \), with \( K^* \) finite.

If \( K \subseteq L \) is \( n \)-Kummer, \( (L^*/n \cap K^*)/K^* \cong G(L/K) \), a finite group.
Thus, the subgroup corresponding to the extension \( K \subseteq L \) is finite.
Conversely, if \( K^* \leq S \leq K^* \) is finite, and \( a_1, \ldots, a_m \in K^* \) are the representatives of the classes in \( S/K^* \), then
\[
K \subseteq K(\sqrt[n]{S}) = K(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_m})
\]
is an \( n \)-Kummer extension (see 3.12). Clearly, these correspondences are inclusion preserving. They are also inverse to each other, as we now prove.

Let \( K \subseteq L \) be \( n \)-Kummer. Its associated subgroup is \( S = L^*/n \cap K^* \).
To \( S/K^* \) corresponds \( K(\sqrt[n]{S}) \), equal to \( L \) (by 3.12).

Conversely, let \( K^* \leq S \leq K^* \), where \( \Delta := S/K^* \) is a finite group.
Its associated \( n \)-Kummer extension is \( K \subseteq K(\sqrt[n]{S}) := L \). Let \( (L^*/n \cap K^*)/K^* = : \Delta' \). We have to prove that \( \Delta = \Delta' \Leftrightarrow S = L^*/n \cap K^* \).
Evidently, \( S \subseteq L^*/n \cap K^* \), so \( \Delta \subseteq \Delta' \). For the other inclusion, define
\[
p : G(L/K) \times \Delta \to U_n, p(\sigma, xK^*) := \sigma(\sqrt[n]{x})/\sqrt[n]{x}, \forall \sigma \in G(L/K),
\]
\[
\forall xK^* \in \Delta.
\]

As in a), one shows that \( p \) is correctly defined and bilinear. It is nondegenerate: if \( \sigma \in G(L/K) \) and \( p(\sigma, xK^*) = 1, \forall xK^* \in \Delta \), then \( \sigma(\sqrt[n]{x}) = \sqrt[n]{x}, \forall x \in S \). But \( L = K(\sqrt[n]{S}) \), so \( \sigma = \text{id} \). If \( xK^* \in S/K^* \) with \( p(\sigma, xK^*) = 1, \forall \sigma \in G(L/K) \), then \( \sqrt[n]{x} \) is fixed by any \( \sigma \in G(L/K) \), so \( \sqrt[n]{x} \in K \) and \( x \in K^* \). Consequently, \( xK^* = 1K^* \).

From 3.15.c), \( \Delta \cong \text{Hom}(G(L/K), U_n) \) via the isomorphism \( bK^* \mapsto \chi_b, \chi_b(\sigma) = (\sigma(\sqrt[n]{b})/\sqrt[n]{b}, \forall bK^* \in S/K^* \), \( \forall \sigma \in G(L/K) \). But \( \Delta' \cong \text{Hom}(G(L/K), U_n) \), by a). Since \( \Delta \subseteq \Delta' \) and the groups are finite, they are equal. \( \square \)

Using the same methods, an analogous result can be obtained for Abelian extensions of exponent \( p \) of a field \( K \) having characteristic \( p > 0 \). For any extension \( K \subseteq L \), we use the following notations:
$L^+$ is the additive group $(L, +)$;
$\mathcal{P} : \Omega^+ \rightarrow \Omega^+$ is the group homomorphism given by $\mathcal{P}(x) = x^p - x$, $\forall x \in \Omega$;
$tp(L^+ / K^+)$ denotes the subgroup in $K^+$ obtained by the $p$-th powers of elements of $L^+$.

The additive homomorphism $\mathcal{P}$ plays the role of the multiplicative homomorphism $x \mapsto x^n$ in the multiplicative theory. For any $a \in K^+$, $\mathcal{P}^{-1}(a)$ denotes any root in $\Omega$ of $X^p - X - a$ (3.8 ensures that, if $b$ is a root, the roots of $X^p - X - a$ are $b, b + 1, \ldots, b + (p - 1)$).

The following results are the “additive” versions of 3.12, 3.14, 3.16. The proofs are proposed as an exercise.

**3.17 Proposition.** Let $K$ be a field of characteristic $p > 0$ and let $K \subseteq L$ be a finite extension. Then $K \subseteq L$ is Abelian of exponent $p$ if and only if there exist $a_1, \ldots, a_m \in K^+ \setminus \mathcal{P}(K^+)$ such that $L = K(\mathcal{P}^{-1}(a_1), \ldots, \mathcal{P}^{-1}(a_m))$. For such an extension, $L = K(\mathcal{P}^{-1}(S))$, where $S := \mathcal{P}(L^+) \cap K^+$, a subgroup in $K^+$.

**3.18 Proposition.** Let $K$ be a field of characteristic $p > 0$ and let $K \subseteq L$ be a finite Abelian extension of exponent $p$. Then the canonical group homomorphism

$\psi : t_p(L^+ / K^+) \rightarrow K^+ / \mathcal{P}(K^+)$, $\psi(x + K^+) = \mathcal{P}(x) + \mathcal{P}(K^+)$, $\forall x \in L^+$ cu $\mathcal{P}(x) \in K^+$,

is injective. The homomorphism $\psi$ induces an isomorphism

$t_p(L^+ / K^+) \cong \text{Im}\psi = (\mathcal{P}(L^+) \cap K^+)) / \mathcal{P}(K^+)$, a subgroup in $K^+ / \mathcal{P}(K^+)$.

**3.19 Theorem.** (Additive Kummer Theory) Let $K$ be a field of characteristic $p > 0$ and let $\mathbb{F}_p$ be its prime subfield. $\mathbb{F}_p^+$ is a cyclic group of order $p$.

a) If $K \subseteq L$ is Abelian of exponent $p$, then there exists a canonical isomorphism

$t_p(L^+ / K^+) \cong \text{Hom}(G(L/K), \mathbb{F}_p^+)$,
VI. Applications of Galois Theory

\[ a + K^+ \mapsto \chi_a, \text{ where } \chi_a(\sigma) = \sigma(a) - a, \]
\[ \forall \ a + K^+ \in \text{tr}(L^+/K^+), \ \forall \sigma \in G(L/K). \]
We also have a canonical isomorphism

\[ (\mathcal{P}(L^+) \cap K^+)/\mathcal{P}(K^+) \cong \text{Hom}(G(L/K), \mathbb{F}_p^+), \]
\[ b + \mathcal{P}(K^+) \mapsto \chi_b, \text{ where } \chi_b(\sigma) = \sigma(\mathcal{P}^{-1}(b)) - \mathcal{P}^{-1}(b), \]
\[ \forall \ b + \mathcal{P}(K^+) \in (\mathcal{P}(L^+) \cap K^+)/\mathcal{P}(K^+), \ \forall \sigma \in G(L/K). \]

Therefore \( G(L/K) \) is isomorphic to the finite subgroup \((\mathcal{P}(L^+) \cap K^+)/\mathcal{P}(K^+)\) of \( K^+/\mathcal{P}(K^+) \).

b) The set of Abelian extensions of exponent \( p \) of \( K \) included in \( \Omega \) is in a bijective inclusion preserving correspondence with the finite subgroups of \( K^+/\mathcal{P}(K^+) \):
the extension \( K \subseteq L \) corresponds to the subgroup \((\mathcal{P}(L^+) \cap K^+)/\mathcal{P}(K^+)\);
the subgroup \( S/K^+ \), with \( \mathcal{P}(K^+) \leq S \leq K^+ \), corresponds to the extension \( K \subseteq K(\mathcal{P}^{-1}(S)) \).

The degree of the extension is equal to the order of the corresponding subgroup.

\[ \square \]

Exercises

1. Let \( \omega \) be a primitive 12\(^{th}\) root of unity. Determine the Galois group of \( \mathbb{Q} \subseteq \mathbb{Q}(\omega, \sqrt[12]{2}) \) and all its intermediate fields.
2. For any \( n \in \mathbb{N}^* \), let \( \zeta_n \) be a primitive \( n \)th root of unity. Prove that, if \( m, n \in \mathbb{N}, (m, n) = 1 \), then \( \mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q} \).
3. Let \( p_1, \ldots, p_n \) be distinct primes. Then \( \mathbb{Q} \subseteq \mathbb{Q}(\sqrt[1]{p_1}, \ldots, \sqrt[20]{p_n}) \) has degree \( 2^n \). Find its Galois group.
4. Determine the degree of \( \mathbb{Q} \subseteq \mathbb{Q}\{\sqrt{d} \mid 1 \leq d \leq 20\} \). Generalization.
5. Let $K = \mathbb{C}(X, Y)$ (the fraction field of the polynomial ring $\mathbb{C}[X, Y]$). Let $L = K(\sqrt[n]{XY}, \sqrt[n]{Y^3})$, where, $\forall t \in K$, $\sqrt[n]{t}$ denotes a root of $T^n - t$ (in some extension of $K$). Prove that $K \subseteq L$ is a 4-Kummer extension and determine its degree and all its intermediate fields.

6. Construct a 6-Kummer extension of degree 18 of $\mathbb{C}(X, Y)$.

VI.4 Solvability by radicals

The classical problem of “solving by radicals an algebraic equation” requires expressing the roots of a given polynomial as a function of the coefficients of the polynomial, using only the four arithmetic operations and radicals (of any order). A typical example is the formula for the roots of the equation of second degree (known to the Babylonians, about 1900-1600 B.C.). In the 16th century, the formulas for the roots of any polynomial equation of degree 3 or 4 were discovered. These successes led many mathematicians of that time to think that such formulas exist for algebraic equations of any degree. For instance, Euler\textsuperscript{24}, around 1749, believed that “the expressions for the roots contain no other operations than radicals, except for the four vulgar operations, and one cannot sustain that transcendental operations might be involved”. These ideas were proven to be false at the beginning of the 19th century by Paolo Ruffini, who proved the impossibility of solving the “general” equation of degree 5 by radicals. Independently from Ruffini (whose intricate proof failed to convince many fellow mathematicians), the Norwegian mathematician Niels

\textsuperscript{24} Leonhard Euler (1707-1783), famous Swiss mathematician.
Henrik Abel gave in 1824 (at the age of 22) a clear and rigorous proof of this impossibility. In 1829, two months before his death, Abel published a memoir in which he describes a class of polynomials solvable by radicals (namely the polynomials whose Galois group is commutative). The commutative groups are called today Abelian, in his honor. In 1830, Evariste Galois, unaware of Abel's results, creates the notion of group (of permutations) and formulates a general criterion of solvability by radicals of a polynomial equation, using what we call today the Galois group of the polynomial. The ideas and results of Galois had a decisive contribution to the development of Algebra.

4.1 Definition. A field extension $K \subseteq L$ is called a radical extension if there exists a finite tower of intermediate fields $K = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_m = L$, such that, for any $i$, $0 \leq i < m$, there exists $a_i \in K_{i+1}$ and $n_i \in \mathbb{N}^*$, with $K_{i+1} = K_i(a_i)$ and $a_i^{n_i} \in K_i$. If $n_1 = \ldots = n_m = n$, $K \subseteq L$ is called an $n$-radical extension.

Let $f \in K[X]$ and let $E$ be the splitting field of $f$ over $K$. The Galois group $G(E/K)$ is called the Galois group of the polynomial $f$ over $K$. We say that the polynomial $f$ is solvable by radicals over $K$ (or that the equation $f = 0$ is solvable by radicals over $K$) if there exists a radical extension $L$ that includes the splitting field of $f$ over $K$.

These definitions articulate rigorously the notion of “formula using radicals and the four operations”. For instance, if $a, b \in \mathbb{Q}$,

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{a}) \subseteq \mathbb{Q}(\sqrt{a})(\sqrt[3]{b - 3a}) = L,$$

is a 6-radical extension. Any element of $L$ is expressed in a basis of $L$ over $\mathbb{Q}$ (for instance $1, \alpha, \beta, \beta^2, \alpha \beta, \alpha \beta^2$, where $\alpha = \sqrt{a}$, $\beta = \sqrt[3]{b - 3a}$) as an “expression using radicals and the four operations”.

4.2 Examples. a) The equation \( aX^2 + bX + c = 0 \), where \( a, b, c \in \mathbb{C}, a \neq 0 \), is solvable by radicals over \( \mathbb{Q}(a,b,c) \). Indeed, the formula for the roots, \( x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \), shows that the roots \( x_i \) are in the 2-radical extension \( \mathbb{Q}(a,b,c) \subseteq \mathbb{Q}(a,b,c)(\sqrt{b^2 - 4ac}) \). Note that, if \( a, b, c \notin \mathbb{Q} \), then the equation may not be solvable over \( \mathbb{Q} \).

b) Any polynomial \( f \in \mathbb{R}[X] \) is solvable by radicals over \( \mathbb{R} \). Indeed the extension \( \mathbb{R} \subseteq \mathbb{C} \) is radical (why?) and \( \mathbb{C} \) is algebraically closed, so it includes the splitting field of \( f \) over \( \mathbb{R} \).

Historically, solvability by radicals means “solvability by radicals over \( \mathbb{Q} \) of polynomials in \( \mathbb{Q}[X] \)”.

We want to prove the following result (Galois' criterion of solvability by radicals): If \( \text{char } K = 0 \), a polynomial is solvable by radicals over \( K \) if and only if its Galois group over \( K \) is solvable.

4.3 Remarks. a) If \( K \subseteq L \) is a radical extension, as in the definition above, then \( K \subseteq L \) is \( n \)-radical, where \( n = \text{LCM}(n_1, \ldots, n_m) \).

b) The \( (n-) \)radical extensions are transitive: If \( K \subseteq L \text{ and } L \subseteq M \) are \( (n-) \)radical, then \( K \subseteq M \) is \( (n-) \)radical. The converse is false: if \( K \subseteq M \) is \( n \)-radical and \( K \subseteq L \subseteq M \), then \( L \subseteq M \) is \( n \)-radical, but \( K \subseteq L \) is not necessarily radical (see example 5.6).

c) If \( K \subseteq L \) is such that \( L = K(x_1, \ldots, x_n) \), where, for each \( i \), \( x_i^m \in K \) for some \( m \in \mathbb{N}^* \), then \( K \subseteq L \) is \( m \)-radical.

4.4 Proposition. The normal closure of an \( n \)-radical extension is an \( n \)-radical extension.

Proof. Let \( K \subseteq L \) be \( n \)-radical and let \( N \) be its normal closure. There exists a tower \( K = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_m = L \) and \( a_i \in K_{i+1} \), \( 0 \leq i < m \), with \( a_i^n \in K_i \) and \( K_{i+1} = K_i(a_i) \).
We prove by induction on $m$ that $K \subseteq N$ is radical. If $m = 1$, then $L = K(a)$, with $a^n = b \in K$, and $N = K(x_1, \ldots, x_r)$, where $x_1, \ldots, x_r$ are the roots of $g := \text{Irr}(a, K)$. Since $g \mid X^n - b$, we have $x_i^n = b$, $1 \leq i \leq r$, so $K \subseteq N$ is $n$-radical.

If $m > 1$, let $M$ be the normal closure of $K \subseteq K_{m-1}$. By induction, $K \subseteq M$ is $n$-radical. It is enough to prove that $M \subseteq N$ is $n$-radical (and then apply the transitivity of $n$-radical extensions). Note that $M$ is the splitting field over $K$ of the family of polynomials $\{\text{Irr}(a_i, K) \mid 1 \leq i \leq m - 1\}$, and $N$ is the splitting field over $K$ of the family of polynomials $\{\text{Irr}(a_i, K) \mid 1 \leq i \leq m\}$. So, $N = M(c_1, \ldots, c_t)$, where $c_1, \ldots, c_t$ are the roots of $h := \text{Irr}(a_m, K)$. Let $b \in K_{m-1}$ such that $a_m^n = b$. For a fixed $i$, $1 \leq i \leq t$, $K(a_m)$ and $K(c_i)$ are $K$-isomorphic via an isomorphism that takes $a_m$ in $c_i$. Extend this homomorphism to a $K$-automorphism $\sigma_i : \Omega \rightarrow \Omega$ such that $\sigma_i(a_m) = c_i$ (see IV.2.18 and IV.2.19). Thus, $c_i^n = \sigma_i(a_m^n) = \sigma_i(b)$. Since $K \subseteq M$ is normal, $\sigma_i(b) \in M$. So, $c_i^n \in M$, $\forall i$, $1 \leq i \leq t$, and therefore $M \subseteq M(c_1, \ldots, c_t) = N$ is $n$-radical.

Thus, if an extension $K \subseteq L$ is contained in a radical extension $K \subseteq E$, we may assume (taking the normal closure) that $K \subseteq E$ is radical and normal.

The following proof of Galois' characterization of solvability by radicals uses basic facts on solvable groups (any subgroup and any factor group of a solvable group are solvable; any finite solvable group has a normal series with the factors cyclic groups), which can be found in the Appendices.

4.5 Theorem. (Galois) Let $K$ be a field of characteristic 0 and let $f \in K[X]$. The polynomial $f$ is solvable by radicals over $K$ if and only if the Galois group of $f$ over $K$ is solvable.

Proof. The theorem follows by setting $L =$ the splitting field of $f$ over $K$ in the statement below:
If char $K = 0$ and $K \subseteq L$ is a normal extension, there exists a radical extension $K \subseteq M$ with $L \subseteq M$ if and only if $G(L/K)$ is a solvable group.

Let us prove this claim. Let $K \subseteq L \subseteq M$ with $L/K$ normal and $M/K$ $n$-radical (for some $n \in \mathbb{N}$). Let us show that $G(L/K)$ is solvable. By 4.4 we may suppose that $M/K$ is normal; since char $K = 0$, it is separable (thus Galois) and $n$-radical. By Galois Theory, $G(M/L) \triangleleft G(M/K)$ and $G(L/K) \cong G(M/K)/G(M/L)$. If we show that $G(M/K)$ is solvable, $G(L/K)$ will be solvable, as a factor of a solvable group.

Let us show that $G(M/K)$ has a solvable series (a normal series with the factors Abelian groups). We use the characterization of the extensions of the form $K \subseteq K(\sqrt[n]{a})$ (prop. 3.3), extensions that come up in the definition of a radical extension. In order to apply this proposition, we need that $\omega \in K$ ($\omega$ is a primitive $n$th root of unity in $\Omega$; $\omega$ exists, since char $K = 0$). Consider the extension $K \subseteq M(\omega)$. Since $K \subseteq M$ and $M \subseteq M(\omega)$ are $n$-radical, $K \subseteq M(\omega)$ is $n$-radical. In the tower of extensions

$$K \subseteq K(\omega) \subseteq M(\omega),$$

$K \subseteq K(\omega)$ is an Abelian extension (by 3.1) and $K(\omega) \subseteq M(\omega)$ is $n$-radical. Thus, there exists a sequence

$$K(\omega) = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_m = M(\omega)$$

and $a_i \in K_{i+1}$, $0 \leq i < m$, with $a_i^n \in K_i$ and $K_{i+1} = K_i(a_i)$. Because $\omega \in K_i$, $\forall i$, 3.3 implies that $K_i \subseteq K_{i+1}$ is cyclic. Let $H_i = G(M(\omega)/K_i)$. Thus, the group $G := G(M(\omega)/K)$ has the following series

$$1 = H_m \subseteq H_{m-1} \subseteq \ldots \subseteq H_1 \subseteq H_0 \subseteq G.$$

We claim that this is a solvable series. $K_{i-1} \subseteq K_i$ is normal, $\forall i$, $1 \leq i \leq m$, so $H_i \triangleleft H_{i-1}$ and $H_{i-1}/H_i \cong G(K_i/K_{i-1})$ is cyclic; also, $K \subseteq K(\omega) = K_0$ is normal, so $H_0 \triangleleft G$ and $G/H_0 \cong G(K(\omega)/K)$ is Abelian. Therefore $G$ is solvable.
Conversely, suppose $G(L/K)$ is a solvable group. There exists a series of subgroups

$$1 = H_m \subseteq H_{m-1} \subseteq \ldots \subseteq H_1 \subseteq H_0 \subseteq G,$$

with $H_i < H_{i-1}$ and $H_{i-1}/H_i$ cyclic, $\forall i$, $1 \leq i \leq m$. Let $K_i$ be the fixed field of $H_i$. The extension $K \subseteq L$ is Galois, so $K_i \subseteq K$ is Galois, $\forall i$, $1 \leq i \leq m$, and $H_{i-1}/H_i \cong G(K_i/K_{i-1})$ is cyclic. Let $\omega$ be a primitive $n$th root of unity in $\Omega$. The extension $K_{i-1}(\omega) \subseteq K_i(\omega)$ is Galois and $G(K_i(\omega)/K_{i-1}(\omega))$ is isomorphic to a subgroup of $G(K_i/K_{i-1})$ (we used V.4.8), so it is cyclic. By 3.3, there exists $a_i \in K_{i+1}(\omega)$, with $a_i^n \in K_i(\omega)$ and $K_{i+1}(\omega) = K_i(\omega)(a_i)$, $\forall i$, $0 \leq i < m$. So, $K(\omega) \subseteq L(\omega)$ is radical; since $K \subseteq K(\omega)$ is radical, $K \subseteq L(\omega)$ is radical (and clearly $L \subseteq K(\omega)$).

**4.6 Example** (The general equation of degree $n$). Let $K$ be a field, let $n \in \mathbb{N}^*$ and let $F$ be the field $K(X_1, \ldots, X_n)$ of rational fractions in $n$ indeterminates. Consider $S := K(s_1, \ldots, s_n)$, the subfield of symmetric rational fractions in $F$ (where $s_1, \ldots, s_n$ are the fundamental symmetric polynomials in $X_1, \ldots, X_n$). The following polynomial in the indeterminate $X$

$$P := (X - X_1)(X - X_2) \ldots (X - X_n) = X^n - s_1 X^{n-1} + s_2 X^{n-2} - \ldots + (-1)^n s_n$$

has coefficients in $S$ and its roots are $X_1, \ldots, X_n \in F$. The polynomial $P$ is called the **generic polynomial of degree $n$**; the equation $P(x) = 0$ is called the **general equation of degree $n$**. The name comes from the fact that $s_1, \ldots, s_n$ are algebraically independent over $K$ (see IV.4.3), so they behave like indeterminates.

Since $F = S(X_1, \ldots, X_n)$, $F$ is the splitting field of $P$ over $S$. The roots of $P$ are distinct, so the extension $S \subseteq F$ is separable and normal. The Galois group $G := G(F/S)$ is isomorphic to $S_n$, the symmetric group on $n$ elements. To prove this, we remark that any $\sigma \in S_n$ induces a $K$-algebra homomorphism $\sigma^* : K[X_1, \ldots, X_n] \to K[X_1, \ldots, X_n]$ such that $\sigma^*(X_i) = X_{\sigma(i)}$. It is immediate that $(\sigma \tau)^* = \sigma^* \tau^*$ and $\text{id}^* = \text{id}$, $\forall \sigma, \tau \in S_n$. Thus, $\sigma^*$ is an isomorphism and its inverse is $(\sigma^{-1})^*$. Passing to the fraction field $K(X_1, \ldots, X_n) = F$, we obtain an isomorphism
(denoted also by $\sigma^*$) $\sigma^* : F \to F$. Since $\sigma^*(s_i) = s_i$, $1 \leq i \leq n$, $\sigma^*$ fixes the subfield $S = K(s_1, ..., s_n)$. In other words, $\sigma^* \in G(F/S)$. Thus, $\sigma \mapsto \sigma^*$ defines a homomorphism (clearly injective) from $S_n$ to $G$. On the other hand, $G$ has at most $n!$ elements, as the Galois group of the splitting field of $f$ a polynomial of degree $n$ (see Example IV.2.16). $S_n$ has $n!$ elements, so $S_n \cong G$. Also note that $P$ is irreducible over $S$ (the root $X_1$ of $P$ has $n = \deg P$ conjugates $X_1, ..., X_n$).

In conclusion, given a field $K$, we constructed a (transcendental...) extension $S$ of $K$ such that there exists a Galois extension $S \subseteq F$ whose Galois group is $S_n$. More generally, given a field $K$ and a finite group $G$, does there exist a Galois extension $K \subseteq L$ whose Galois group is $G$? This problem is extremely difficult in the general case. For example, a theorem of I.R. Shafarevitch (proven around 1954 in several papers totaling more than 100 pages) has as a consequence that, for any solvable finite group $G$, there exists a Galois extension of $\mathbb{Q}$, whose Galois group is $G$. The results of this type are known as Constructive Galois Theory.

Back to our example, recalling that $S_n$ is solvable for any $n \leq 4$ implies the solvability of the extension $S \subseteq F$ (if char $K = 0$). In other words, there exists formulas that express by radicals the roots of $P$ if $n \leq 4$. If $a_1, ..., a_n \in K$, there exists a unique $K$-algebra homomorphism $\varphi : K[s_1, ..., s_n] \to K$ such that $\varphi(s_i) = a_i$, $1 \leq i \leq n$ ($K[s_1, ..., s_n]$ is isomorphic to the $K$-algebra of polynomials in $n$ indeterminates and use the universality property of this $K$-algebra). So, replacing the “indeterminates” $s_1, ..., s_n$ with $a_1, ..., a_n$, we obtain formulas that solve by radicals any polynomial equation of degree $n \leq 4$ with coefficients in $K$.

For $n \geq 5$, $S_n$ is not solvable, so the general equation of degree $n$ is not solvable by radicals over $K$ (this is the Abel-Ruffini Theorem). But there exist particular equations of degree $\geq 5$, solvable by radicals (like $X^5 = 0$ ...).
VI.5 Discriminants, resultants

The notion of discriminant of a polynomial of degree 2 is well-known. If \( g = aX^2 + bX + c \in \mathbb{R}[X] \), \( g \) has a double root if and only if the discriminant \( D = b^2 - 4ac \) is zero. If \( x_1, x_2 \) are the roots of \( g \), then \( D = a^2(x_1 - x_2)^2 \) (easy exercise). This suggests the following general definition:

5.1 Definition. a) Let \( K \) be a field, let \( g \in K[X] \) be a polynomial of degree \( n \) and let \( x_1, \ldots, x_n \) be the roots of \( g \) (in some algebraic closure \( \Omega \) of \( K \)). The discriminant of \( g \) is

\[
D(g) := a^{2n-2} \prod_{i<j} (x_j - x_i)^2,
\]

where \( a \) is the leading coefficient of \( g \).

Note that \( D(g) \) is a symmetric polynomial in the roots \( x_1, \ldots, x_n \) of \( g \). This implies that \( D(g) \in K \) and that \( D(g) \) is independent on the labeling of the roots of \( g \). Moreover, \( g \) has a double root if and only if \( D(g) = 0 \).

It is useful to define also \( \Delta := \prod_{i<j} (x_j - x_i) \in \Omega \). In general, \( \Delta \not\in K \) and relabeling the roots \( x_1, \ldots, x_n \) may change \( \Delta \) in \(-\Delta\). Thus, \( \Delta \) is defined up to a sign.

b) If \( K \subseteq L \) is a field extension and \( x \in L \) is algebraic over \( K \), then the discriminant of the element \( x \) over \( K \) is \( D_K(x) := D(\text{Irr}(x, K)) \).

We use the discriminant to obtain data on the Galois group \( G \) of the polynomial \( g \), assuming \( g \) has only simple roots (it is separable). To this end, we see the elements of the Galois group as permutations of the roots.

5.2 Definition. Let \( L \) be the splitting field of \( g \) over \( K \) and let \( x_1, \ldots, x_n \) be its (distinct) roots in \( L \). For any \( \sigma \in G = G(L/K) \), let \( \varphi(\sigma) \in S_n \) be the permutation given by \( \varphi(\sigma)(i) = j \iff \sigma(x_i) = x_j, \forall i, j \in \{1, \ldots, n\} \). The mapping \( \varphi : G \to S_n \) is an injective group homo-
morphism (see IV.3.15,c)) and identifies \( G \) with a subgroup of \( S_n \). If we reorder the roots (say \( x'_1, \ldots, x'_n \), where \( x'_i = x_{\tau(i)} \) for some \( \tau \in S_n \)) then \( \varphi \) is replaced by \( \varphi' : G \to S_n \). We have \( \varphi'(\sigma)(i) = j \iff \sigma(x'_i) = x'_j \iff \varphi(\sigma)(\tau(i)) = \tau(j) \); thus \( \varphi'(\sigma) = \tau^{-1} \circ \varphi(\sigma) \circ \tau \).

This shows that a reordering of the roots leads to replacing \( \varphi(\sigma) \) with a conjugate permutation \( \tau^{-1} \varphi(\sigma) \tau \). Thus, if a permutation (a subgroup of permutations) has a property that is invariant under any conjugation\(^{25}\), then this property can be transferred to the corresponding element in \( G(L/K) \) (respectively to a subgroup in \( G(L/K) \)). Thus, we define:

Let \( \sigma \in G(L/K) \). Define the signature of \( \sigma \), \( \varepsilon(\sigma) := \varepsilon(\varphi(\sigma)) \), where \( \varepsilon(\varphi(\sigma)) \) denotes the signature of the permutation \( \varphi(\sigma) \in S_n \). We say that \( \sigma \) is even if \( \varepsilon(\sigma) = 1 \iff \varphi(\sigma) \) is an even permutation in \( S_n \). We say that \( \sigma \) is odd if \( \varepsilon(\sigma) = -1 \).

If \( 1 \leq k \leq n \), we say that \( \sigma \) is a cycle of length \( k \) if \( \varphi(\sigma) \) is a cycle of length \( k \).

These definitions are independent on the labeling of the roots of \( g \). Indeed, \( \varepsilon(\tau^{-1} \varphi(\sigma) \tau) = \varepsilon(\varphi(\sigma)) \) (see above); also, the conjugate of a cycle of length \( k \) is also a cycle of length \( k \).

5.3 Proposition. Keep the previous notations and suppose \( \text{char } K \neq 2 \). Let \( g \in K[X] \) be a polynomial without multiple roots and let \( \sigma \in G(L/K) \). Then:

a) \( \sigma(\Delta) = \varepsilon(\sigma) \Delta \). In other words: \( \sigma \in G(L/K) \) is even \( \iff \sigma(\Delta) = \Delta \). Also, \( \sigma \in G(L/K) \) is odd \( \iff \sigma(\Delta) = -\Delta \).

b) All \( \sigma \in G(L/K) \) are even \( \iff \Delta \in K \iff D(g) \in K^2 \).

\(^{25}\) If \( \tau \in S_n \), the conjugation by \( \tau \) is the automorphism \( \kappa_{\tau} : S_n \to S_n \) defined by \( \kappa_{\tau}(\eta) = \tau^{-1} \eta \tau, \forall \eta \in S_n \).
322 VI. Applications of Galois Theory

c) \( \text{Gal}(L/K(\Delta)) = \{ \sigma \in G(L/K) \mid \sigma \text{ is even} \} \). In other words, the Galois correspondences take the intermediate field \( K(\Delta) \) to the subgroup \( G(L/K) \cap \varphi^{-1}(A_n) \) of even elements in \( G(L/K) \).

**Proof.** a) In our hypotheses, \( \Delta \neq 0, \Delta \neq -\Delta \) and \( K \subseteq L \) is a Galois extension. If \( \tau \in S_n \), let \( \tau^* : K[X_1, \ldots, X_n] \to K[X_1, \ldots, X_n] \) denote the unique \( K \)-algebra homomorphism such that \( \tau^*(X_i) = X_{\tau i} \). Let \( d := \prod_{i<j} (X_j - X_i) \in K[X_1, \ldots, X_n] \). We show now that \( \tau^*(d) = \varepsilon(\tau)d \).

For any \( 1 \leq i < j \leq n \), define:

\[
\varepsilon(i, j) := \begin{cases} 1 & \text{if } \tau i < \tau j \\ -1 & \text{if } \tau i > \tau j \end{cases}
\]

Note that

\[
\{(X_j - X_i) \mid 1 \leq i < j \leq n\} = \{(X_{\tau i} - X_{\tau j}) \cdot \varepsilon(i, j) \mid 1 \leq i < j \leq n\},
\]

so

\[
\tau^*(d) = \prod_{i<j} (X_{\tau j} - X_{\tau i}) = \prod_{i<j} \varepsilon(i, j)(X_j - X_i) = \varepsilon(\tau)d.
\]

We have \( \Delta = d(x_1, \ldots, x_n) \). Let \( \sigma \in G(L/K) \). Then:

\[
\sigma(\Delta) = (\varphi(\sigma)^*(d))(x_1, \ldots, x_n) = \varepsilon(\varphi(\sigma))d(x_1, \ldots, x_n) = \varepsilon(\sigma)\Delta.
\]

b) If \( G(L/K) \subseteq A_n \) (i.e. any \( \sigma \in G(L/K) \) is even), then \( \sigma(\Delta) = \Delta \), for any \( \sigma \in G(L/K) \). Since \( K \subseteq L \) is Galois, this implies \( \Delta \in K \), so \( D(\Delta) = \Delta^2 \in K^2 \).

If \( D(\Delta) \in K^2 \), then \( D(\Delta) = a^2 \), for some \( a \in K \). But \( D(\Delta) = \Delta^2 \), so \( \Delta^2 = a^2 \Rightarrow \Delta = a \) or \( \Delta = -a \). Anyway, \( \Delta \in K \), thus, \( \forall \sigma \in G(L/K), \sigma(\Delta) = \Delta \), which means that \( \sigma \) is even.

c) \( G(L/K(\Delta)) = \{ \sigma \in G(L/K) \mid \sigma(\Delta) = \Delta \} = \{ \sigma \in G(L/K) \mid \sigma \text{ is even} \} = G(L/K) \cap \varphi^{-1}(A_n) \). \( \square \)

We need a method to compute the discriminant of a polynomial without knowing its roots.

**5.4 Proposition.** Let \( g = a_0 + a_1X + \ldots + a_{n-1}X^{n-1} + X^n \in K[X] \) and let \( x_1, \ldots, x_n \) be the roots of \( g \) in \( \Omega \). For any \( m \in \mathbb{N} \), let \( t_m := x_1^m + \ldots + x_n^m \in K \). Then
VI.5 Discriminants, resultants

\[ D(g) = \det \begin{bmatrix} t_0 & t_1 & \ldots & t_{n-1} \\ t_1 & t_2 & \ldots & t_n \\ \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_n & \ldots & t_{2n-2} \end{bmatrix} \]

The sums \( t_m \) can be computed by recurrence using the following relations (Newton's identities):

\[ t_0 = n; \quad t_1 = -a_{n-1}; \]
\[ -t_m = a_{n-1} t_{m-1} + a_{n-2} t_{m-2} + \ldots + a_{n-m+1} t_1 + a_{n-m} m, \text{ if } 2 \leq m \leq n; \]
\[ -t_m = a_{n-1} t_{m-1} + a_{n-2} t_{m-2} + \ldots + a_0 t_{m-n}, \quad \text{ if } m > n. \]

**Proof.** \( \Delta = \prod_{i<j} (x_j - x_i) = \det A \), where \( A \) is the Vandermonde matrix:

\[
A = \begin{bmatrix}
1 & x_1 & \ldots & x_1^{n-1} \\
1 & x_2 & \ldots & x_2^{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_n & \ldots & x_n^{n-1}
\end{bmatrix}
\]

The matrix \( A^t A \) is the matrix in the statement, and \( \det(A^t A) = \det(A)^2 = \Delta^2 = D(g) \).

Viète's relations coupled with Newton's identities (see Appendix) yield the formulas for the \( t_m \). \( \square \)

The discriminant can also be computed using the **resultant**, defined below.

**5.5 Example.** Let us compute the discriminant of a polynomial of degree 3, \( X^3 + pX + q \). We have \( n = 3 \), \( a_0 = q \), \( a_1 = p \), \( a_2 = 0 \). We have \( t_0 = 3; \ t_1 = 0. \) Newton's identities yield: \( t_2 = -2p; \ t_3 = -3q; \ t_4 = 2p^2 \). Thus

\[
D(X^3 + pX + q) = \begin{vmatrix} 3 & 0 & -2p \\ 0 & -2p & -3q \\ -2p & -3q & 2p^2 \end{vmatrix} = -4p^3 - 27q^2.
\]
5.6 Example. We want to give an example of polynomial \( g \in \mathbb{Q}[X] \) solvable by radicals over \( \mathbb{Q} \), but such \( L \), the splitting field of \( g \) over \( \mathbb{Q} \), is not a radical extension of \( \mathbb{Q} \). We claim that any irreducible polynomial \( g \in \mathbb{Q}[X] \), of degree 3, such that the splitting field of \( g \) over \( \mathbb{Q} \) has degree 3, has this property. Indeed, a normal extension \( \mathbb{Q} \subseteq \mathbb{L} \) of degree 3 cannot be radical. Assuming \( \mathbb{Q} \subseteq \mathbb{L} \) is radical, the definition implies \( \mathbb{L} = \mathbb{Q}(b) \), where \( b \in \mathbb{L} \) is a root of \( h = X^3 - a \), for some \( a \in \mathbb{Q} \).

Since the extension is normal, all roots of \( h \) are in \( \mathbb{L} \), so \( b \omega \) and \( b \omega^2 \in \mathbb{L} \), where \( \omega \) is a primitive third root of unity. Thus, \( \omega \in \mathbb{L} \), so \( \mathbb{Q} \subseteq \mathbb{Q}(\omega) \subseteq \mathbb{L} \). It follows that \( 2 = [\mathbb{Q}(\omega) : \mathbb{Q}] \) divides 3 = [\( \mathbb{L} : \mathbb{Q} \)], absurd. On the other hand, \( g \) is solvable by radicals over \( \mathbb{Q} \) because its Galois group \( G(\mathbb{L}/\mathbb{Q}) \) is a subgroup of \( S_3 \), which is solvable.

An example of such polynomial is the minimal polynomial of \( \cos(2\pi/9) \) (see exercise V.4.8). More generally, if \( g = X^3 + pX + q \in \mathbb{Q}[X] \) is irreducible and has the discriminant \( D(g) = -4p^3 - 27q^2 \) a square of a rational number, then (by Proposition 5.3) the Galois group of \( g \) is \( A_3 \) (the alternating group on 3 elements, which is cyclic of degree 3) and \( g \) is solvable by radicals. For instance, \( g = X^3 - 3X + 1 \) has \( D(g) = 81 \in \mathbb{Q}^2 \) and is irreducible (\( g \) has no roots in \( \mathbb{Q} \) and has degree 3).

The resultant of two polynomials \( g \) and \( h \in K[X] \) appears in the problem of finding conditions under which \( g \) and \( h \) have a common nonconstant factor (equivalently, they have a common root in the algebraic closure \( \Omega \) of \( K \)).

5.7 Definition. Let \( K \) be a field and let \( g, h \in K[X] \) be nonzero. Assume that the roots of \( g \) are \( x_1, \ldots, x_m \in \Omega \) and the roots of \( h \) are \( y_1, \ldots, y_n \in \Omega \). Then, in \( \Omega[X] \),

\[
g = a(X - x_1) \cdots (X - x_m), \quad h = b(X - y_1) \cdots (X - y_n),
\]

where \( a, b \in K^* \) and \( m, n \in \mathbb{N} \). The resultant of \( g \) and \( h \) is the product

\[
\text{Res}(g, h) := a^n b^m \cdot \prod_{1 \leq i \leq m, 1 \leq j \leq n} (x_i - y_j).
\]
By convention, $\text{Res}(0, g) = \text{Res}(g, 0) = 0$, $\forall g \in K[X]$.

The next proposition shows that $\text{Res}(g, h) \in K$ and provides an algorithm that computes the resultant using divisions with remainder.

5.8 Proposition. Let $g, h \in K[X]$ be nonzero. Then:

a) $\text{Res}(g, h) = 0 \iff g$ and $h$ have a common root in $\Omega \iff \text{GCD}(g, h)$ is a nonconstant polynomial in $K[X]$.

b) Let $g = a(X - x_1) \ldots (X - x_m), \ h = b(X - y_1) \ldots (X - y_n)$, with $a, b \in K^*$ and $m, n \in \mathbb{N}$. The following formulas hold:

\[
\text{Res}(g, h) = a^{\deg(h)} \prod_{1 \leq i \leq m} h(x_i) \quad (R1)
\]

\[
\text{Res}(h, g) = (-1)^{mn} \text{Res}(g, h) \quad (R2)
\]

\[
\text{Res}(g, h) = a^{n - \deg(r)} \text{Res}(g, r) \quad (R3)
\]

(R is the remainder of the division of $h$ by $g$: $h = gq + r$, where $q, r \in K[X], \deg r < \deg g$ or $r = 0$)

\[
\text{Res}(g, b) = b^m, \forall b \in K \text{ a constant polynomial.} \quad (R4)
\]

c) $\text{Res}(g, h) \in K$.

Proof. a) Clear, keeping in mind that the GCD of $g$ and $h$ in $\Omega[X]$ is the same as their GCD in $K[X]$.

b) We have $h(x_i) = b(x_i - y_1) \ldots (x_i - y_n)$, for any $i, 1 \leq i \leq m$; R1 follows by multiplying these relations.

R2 is clear from the definition.

Let $h = gq + r$, where $r = 0$ or $\deg r < m = \deg h$. From R1,

\[
\text{Res}(g, h) = a^n \prod_i h(x_i) = a^n \prod_i (g(x_i)q(x_i) + r(x_i)) = a^n \prod_i r(x_i),
\]

since $g(x_i) = 0, 1 \leq i \leq m$. Bur $\text{Res}(g, r) = a^{\deg(r)} \prod_i r(x_i)$ and R3 follows.

Finally, R1 implies R4.

c) By induction on $\min(\deg g, \deg h)$, using R3 and R4. 

A polynomial $g$ has multiple roots if and only if $g$ and its formal derivative $g'$ have common roots. Thus $D(g) = 0 \iff \text{Res}(g, g') = 0$. In fact, $D(g)$ and $\text{Res}(g, g')$ differ only by $\pm$ the dominant coefficient of $g$. 
5.9 Proposition. Let \( g \in K[X] \) have degree \( m \in \mathbb{N} \) and let \( a \in K^* \) be its dominant coefficient. Then \( D(g) = a^{-1}(-1)^{m(m-1)/2}\text{Res}(g, g') \).

Proof. Let \( g = a(X - x_1)\ldots(X - x_m) \), where \( x_i \in \Omega \). Then

\[
g' = a\sum_i \prod_{j \neq i} (X - x_j).
\]

By R1,

\[
\text{Res}(g, g') = a^{m-1}\prod_i g'(x_i) = a^{m-1}\prod_i (a\prod_{j \neq i} (x_i - x_j)).
\]

For each couple \((i, j), 1 \leq i < j \leq m\), the product above contains \((x_i - x_j)\) and \((x_j - x_i)\). There are \(m(m-1)/2\) such couples, so

\[
\text{Res}(g, g') = a^{2m-2}(-1)^{m(m-1)/2}\prod_{i<j} (x_i - x_j)^2 = a(-1)^{m(m-1)/2}D(g).
\]

The resultant of two polynomials can be also computed by means of a determinant formed with their coefficients (see the exercises). If \( x \) is a root of \( f \), and \( y \) is a root of \( g \), with the help of resultants one can find a polynomial that has the root \( x + y \) (see Exercise 9).

Exercises

1. Let \( K \subseteq L \) be an extension of prime degree \( p \). If the extension is radical, then there exists \( b \in L \) such that \( L = K(b) \) and \( b^p \in K \). Deduce that a normal extension of degree \( p \) of \( \mathbb{Q} \) cannot be radical.

2. Let \( K \) be a field of characteristic 0 and let \( f \in K[X] \) be a polynomial solvable by radicals over \( K \), \( \deg f = n \). Prove that, if \( K \) contains a primitive \( n! \)-th root of unity, then the splitting field of \( f \) over \( K \) is a radical extension of \( K \).

3. Let \( K \) be a field and let \( p = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0 \in K[X] \), such that the characteristic exponent of \( K \) does not divide \( n \). Prove that solving the equation \( p(X) = 0 \) is equivalent to solving an equation in \( Y \) of the form \( Y^n + b_{n-2}Y^{n-2} + \ldots + b_1Y + b_0 \), by setting \( Y = X + a/n \).
4. *(Solving the cubic equation)*\(^{26}\) Let \(K\) be a field whose characteristic is not 2 or 3 and let the equation

\[
f(x) = x^3 + px + q = 0,
\]

where \(p, q \in K\). Prove the following assertions:

a) Setting \(x = u + v\), one obtains the equation

\[
u^3 + v^3 + q + (3uv + p)(u + v) = 0.
\]

b) Requiring \(u^3 + v^3 + q = 0\) and \(3uv + p = 0\), an equation of degree 2 in \(u^3\) is obtained, whose solutions are \(u^3 = \frac{-q \pm \sqrt{q^2 + 4p^3/27}}{2}\). Let \(D = -4p^3 - 27q^2\) be the discriminant of the equation \(f = 0\). Then \(u^3 = -q/2 \pm \sqrt{\gamma}\), where \(\gamma = -D/108\).

c) Let \(A = -q/2 + \sqrt{\gamma}\), \(B = -q/2 - \sqrt{\gamma}\) and let \(\omega\) be a primitive 3rd root of unity. The condition \(uv = -p/3\) implies that the roots of \(f\) are:

\[
\frac{3\sqrt[3]{A} + 3\sqrt[3]{B}}{2}, \quad \omega^{2/3}\sqrt[3]{A} + \omega^{2/3}\sqrt[3]{B}, \quad \omega^{1/3}\sqrt[3]{A} + \omega^{1/3}\sqrt[3]{B}.
\]

5. *(Solving the quartic equation)*\(^{27}\) Let \(K\) be a field whose characteristic is not 2 or 3 and let the equation

\[
f(x) = x^4 + px^2 + qx + r = 0,
\]

where \(p, q, r \in K\). Verify the details in the following steps, leading to a solution of this equation:

a) Solve the equation if \(q = 0\).

b) Let \(q \neq 0\). Then \(f(x) = 0\) is rewritten:

\[
\left(x^2 + \frac{p}{2}\right)^2 = -qx - r + \frac{p^2}{4}
\]

c) Let \(u \in K\). Then any solution \(x\) of the equation satisfies

\[
26\text{ The formula was obtained around 1515, by Scipione del Ferro (1465 - 1526), but was not published. Niccolò Fontana “Tartaglia” (1500-1557) rediscovered the formula in 1535 and communicated it to Girolamo Cardano (1501-1576), who published it in his book “Ars Magna, sive de regulis algebraicis”.}

\[
27\text{ The formula was found by Lodovico Ferrari (1522-1565), a student of Cardano.}
\]
VI. Applications of Galois Theory

\[
\left(x^2 + \frac{p}{2} + u\right)^2 = -qx - r + \frac{p^2}{4} + u^2 + 2ux^2 + pu
\]

The right hand side is a perfect square of some polynomial in \(x\) if and only if

\[
-qx - r + \frac{p^2}{4} + u^2 + 2ux^2 + pu = \left(x\sqrt{2u - \frac{q}{2\sqrt{2u}}}\right)^2
\]

d) The last equality is equivalent to

\[
8u^3 + 8pu^2 + \left(2p^2 - 8r\right)u - q^2 = 0.
\]

(called the cubic resolvent of Ferrari).

e) Let \(u\) be a solution of the cubic resolvent. Then the 4 solutions of \(f = 0\) are:

\[
x = \varepsilon \sqrt{\frac{u}{2}} \pm \sqrt{\frac{-u}{2} - \frac{p}{2} - \frac{\varepsilon q}{2\sqrt{2u}}}, \text{ where } \varepsilon \in \{-1, 1\}.
\]

6. Let \(g = X^3 + pX + q \in \mathbb{Q}[X]\) be irreducible, and let \(\alpha \in \mathbb{R}\) be a root of \(g\). Prove that, if \(\Delta = -4p^3 - 27q^2\) is the discriminant of \(g\), then \(\mathbb{Q}(\alpha, \sqrt[3]{\Delta})\) is the splitting field of \(g\) over \(\mathbb{Q}\). Deduce that \(\mathbb{Q} \subseteq \mathbb{Q}(\alpha)\) is normal if and only if \(\Delta\) is the square of a rational number.

7. Let \(g = X^3 + X + 3 \in \mathbb{Q}[X]\). Show that \(g\) is irreducible. Let \(\alpha\) be the real root of \(g\) and let \(K\) be the splitting field of \(g\) over \(\mathbb{Q}\). Show that \(K \cap \mathbb{Q}(\alpha, \sqrt[3]{d}) = \mathbb{Q}(\alpha)\), for any \(d \in \mathbb{Z}\) a positive squarefree integer.

8. (The resultant as a determinant) Let \(R\) be a domain, let \(a, b \in R\) and let \(m, n \in \mathbb{N}^*\). Consider the polynomials in \(X\), with coefficients in \(R[X_1, \ldots, X_m, Y_1, \ldots, Y_n]\):

\[
f := a \prod_{1 \leq i \leq m} (X - X_i) = \sum_{0 \leq j \leq m} a_j X^j
\]

\[
g := b \prod_{1 \leq i \leq n} (X - Y_i) = \sum_{0 \leq j \leq n} b_j X^j.
\]

Let \(D\) the square matrix in \(M_{m+n}(R[X_1, \ldots, X_m, Y_1, \ldots, Y_n])\),
There are \( n \) columns of \( a \)'s and \( m \) columns of \( b \)'s (the empty places contain 0's). The point is to prove that \( \text{Res}(f, g) = \det D \).

a) Let \( M \) be the Vandermonde matrix of dimension \( m + n \),

\[
M = \begin{bmatrix}
Y_1^{m+n-1} & \cdots & \cdots & Y_1 & 1 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
Y_n^{m+n-1} & \cdots & \cdots & Y_n & 1 \\
X_1^{m+n-1} & \cdots & \cdots & X_1 & 1 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
X_m^{m+n-1} & \cdots & \cdots & X_m & 1
\end{bmatrix}
\]

Computing \( \det(MD) \) in two ways, show that \( \det D = \text{Res}(f, g) \)

b) Prove that \( \text{Res}(f, g) \in R[a_0, \ldots, a_m, b_0, \ldots, b_n] \).

9. Let \( K \) be a field and let \( x, y \) be algebraic elements over \( K \); let \( f = \text{Irr}(x, K) \) and let \( g = \text{Irr}(y, K) \). Let \( Z \) be an indeterminate and let \( g_Z := g(Z + X) \in K[Z][X] \). View \( f \) and \( g_Z \) as polynomials in \( X \) (with coefficients in \( K[Z] \)), and let \( h := \text{Res}(f, g_Z) \in K[Z] \). Prove that \( x + y \) is a root of \( h \). More precisely, the roots of \( h \) are \( x_i + y_j, \ 1 \leq i \leq n, 1 \leq j \leq m \), where \( \{x_1, \ldots, x_n\} \) are the roots of \( f \) and \( \{y_1, \ldots, y_m\} \) are the roots of \( g \).

10. Let \( C \) be the curve in \( \mathbb{R}^2 \) given by the parametric representation

\[
\begin{cases}
x = x(t) = t^2 + t, \ t \in \mathbb{R} \\
y = y(t) = t^2 - t + 1
\end{cases}
\]

This means that \( C = \{(x, y) \in \mathbb{R}^2 \mid \exists t \in \mathbb{R} \} \).
such that \( x = x(t), \ y = y(t) \}. Using resultants, find an implicit representation of the curve (i.e., a polynomial \( g \in \mathbb{R}[X, Y] \) such that: \( \forall t \in \mathbb{R}, \ g(x(t), \ y(t)) = 0 \); and, conversely, \( g(x, \ y) = 0 \) implies that \( \exists t \in \mathbb{R} \) such that \( x = x(t), \ y = y(t) \). (Hint. \( \forall (x, y) \in \mathbb{R}^2, \ (x, y) \in C \iff \) the polynomials \( t^2 + t - x \) and \( t^2 - t + 1 - y \) have a common root.)
Appendices

The following appendices contain a part of the background required for reading the book. They include themes less likely to be included in a first course on Algebra and themes that are necessary, but auxiliary to the main ideas in the book.

The reader should feel free to consult the appendices as needed, when reading the main text.

1. Prime ideals and maximal ideals

All rings considered are assumed commutative rings with identity. If $I$ is an ideal of the ring $R$ (i.e. $\forall x, y \in I$ and $\forall r \in R$, we have $x + y \in I$ and $rx \in R$), we denote this by $I \leq R$. An ideal $I$ is called a proper ideal if $I \neq R$. If $I$ and $J$ are ideals in $R$ with $I \subseteq J$, we write this also $I \leq J$. The set of all ideals of a ring is a lattice with respect to inclusion: for any two ideals $I$ and $J$ of $R$, $\inf(I, J) = I \cap J$, $\sup(I, J) = I + J$, where $I + J$ denotes the sum of the ideals $I$ and $J$, $I + J := \{i + j \mid i \in I, j \in J\}$ (the ideal generated by $I \cup J$).

Throughout this section, $R$ denotes a commutative ring with identity.
1.1 Definition. An ideal \( P \) of the ring \( R \) is called a **prime ideal** if \( P \neq R \) and, for any \( x, y \in P \), \( xy \in P \) implies \( x \in P \) or \( y \in P \).

An ideal \( M \) of \( R \) is called a **maximal ideal** if \( M \neq R \) and there exist no proper ideals of \( R \) that strictly include \( M \). In other words, for any \( J \leq R, M \leq J \) implies \( M = J \) or \( J = R \).

1.2 Examples. a) If \( p \in \mathbb{Z} \) is a prime number, then the ideal generated by \( p \) in \( \mathbb{Z} \), denoted \( p\mathbb{Z} \), is a prime ideal in \( \mathbb{Z} \). Conversely, if \( p\mathbb{Z} \) is a prime ideal, then \( p \) is a prime number.

b) An ideal \( I \) is maximal in the ring \( R \) iff \( I \) is a maximal element of the set (ordered by inclusion) of all proper ideals of \( R \). In the ring \( \mathbb{Z} \), any ideal is of the form \( n\mathbb{Z} \), for some \( n \in \mathbb{Z} \). This implies that \( n\mathbb{Z} \) is maximal if and only if \( n \) is a prime number. Indeed, let \( n\mathbb{Z} \) be a maximal ideal. Then, \( \forall m \in \mathbb{Z}, n\mathbb{Z} \subseteq m\mathbb{Z} \) implies \( n\mathbb{Z} = m\mathbb{Z} \) or \( m\mathbb{Z} = \mathbb{Z} \); in other words, \( m|n \) implies \( m \sim n \) or \( m = 1 \). Since \( n\mathbb{Z} \neq \mathbb{Z} \), this means that \( n \) is irreducible, hence it is a prime. The converse is left to the reader.

c) The ring \( R \) is a domain (has no zero divisors) if and only if \((0)\) is a prime ideal.

d) If \( K \) is a field, \((0)\) is its only proper ideal; \((0)\) is also a maximal ideal and a prime ideal.

Here is a useful characterization of maximal (respectively prime) ideals, by means of the factor ring.

1.3 Theorem. Let \( I \) be a proper ideal of \( R \).

a) \( I \) is a prime ideal if and only if the factor ring \( R/I \) is a domain.

b) \( I \) is a maximal ideal if and only if the factor ring \( R/I \) is a field.

**Proof.** a) Let \( I \) be a prime ideal. Let \( \alpha = a + I, \beta = b + I \) (where \( a, b \in R \)) be elements in \( R/I \). If \( \alpha\beta = 0 \in R/I \), then \((a + I)(b + I) = 0 + I \), so \( ab \in I \). Since \( I \) is prime, \( a \in I \) or \( b \in I \), hence \( a + I = \alpha = 0 + I \) or \( b + I = \beta = 0 + I \). Therefore, \( R/I \) has no zero divisors. Conversely, assume that \( R/I \) is a domain and let \( a, b \in R \) with \( ab \in I \). This implies
that \((a + I)(b + I) = 0 + I\), so \(a + I = 0 + I\) or \(b + I = 0 + I\). Thus, \(a \in I\) or \(b \in I\).

\(b)\) Suppose \(I\) is a maximal ideal in \(R\). We want to show that any nonzero element of \(R/I\) is invertible. Let \(\alpha = a + I \neq 0 + I\), hence \(a \notin I\). Then the ideal generated by \(I\) and \(a, I + Ra\), includes strictly \(I\); since \(I\) is maximal, \(I + Ra = R\). In particular, \(1 \in R\) is written \(i + ra\), for some \(i \in I\) and \(r \in R\). Thus, \(1 + I = (ra + i) + I = ra + I = (r + I)(a + I)\), so \(a + I\) is invertible.

If \(R/I\) is a field and \(J\) is an ideal strictly including \(I\), there exists \(x \in J, x \notin I\). Hence, \(x + I \neq 0 + I\), so \(x + I\) is invertible in \(R/I\). For some \(r \in R\), \(1 + I = (r + I)(x + I)\), whence there exists \(i \in I\) such that \(1 = rx + i\). It follows that \(1 \in J \iff J = R\). \(\Box\)

\textbf{1.4 Corollary.} Any maximal ideal is a prime ideal. \(\Box\)

The converse is false: the ideal \((X)\) of the ring \(\mathbb{Z}[X]\) is prime and not maximal, because the factor ring \(\mathbb{Z}[X]/(X) \cong \mathbb{Z}\) is a domain, but not a field.

If \(R\) is a principal ideal domain and not a field, the prime nonzero ideals are exactly the maximal ideals and they are the ideals generated by irreducible elements.

Krull's Lemma (II.1.20) says that any proper ideal is included in some maximal ideal.

\section{2. Algebras. Polynomial and monoid algebras}

We fix \((R, +, \cdot)\), a commutative ring with identity.
2.1 Definition. An \textit{R-algebra} is a ring \((A, +, \cdot)\) (not necessarily associative or having an identity), which is simultaneously an \(R\)-module, such that, \(\forall r \in R, \forall a, b \in A:\)

\[ r(ab) = (ra)b = a(rb). \]

The \(R\)-algebra \(A\) is called \textit{associative} (respectively \textit{with identity}, \textit{commutative}) if the ring \(A\) has the corresponding property.

Recall that the \textit{center} of the ring \(A\), \(\text{Cen}(A)\), is the set of all elements in \(A\) that commute with any element:

\[ \text{Cen}(A) := \{ a \in A \mid ab = ba, \forall b \in A \} \]

\(\text{Cen}(A)\) is a subring of \(A\) (easy exercise).

We are interested in \(R\)-algebras that are \textit{associative and have an identity}. For this type of algebras there is the following characterization (often taken as the definition):

2.2 Proposition. \(a)\) Let \(A\) be an associative \(R\)-algebra with identity element \(e\). Then the function \(\alpha : R \rightarrow A\) defined by

\[ \alpha(r) := re, \forall r \in R \]

is an identity preserving ring homomorphism and \(\alpha(r)a = a\alpha(r), \forall r \in R, \forall a \in A\) (in other words, \(\alpha(R) \subseteq \text{Cen}(A)\)).

\(b)\) Conversely, if \(A\) is an associative ring with identity, and \(\alpha : R \rightarrow A\) is an identity preserving ring homomorphism such that \(\alpha(R) \subseteq \text{Cen}(A)\), then \(A\) is an \(R\)-algebra if we define the \(R\)-module multiplication by

\[ ra := \alpha(r)a, \forall r \in R, \forall a \in A. \]

Proof. \(a)\) If \(r, s \in R\), then, by definition:

\[ \alpha(r + s) = (r + s)e = re + se = \alpha(r) + \alpha(s) \]

\[ \alpha(r)\alpha(s) = (re)(se) = r(e(se)) = r(se) = (rs)e = \alpha(rs). \]

Also, \(\alpha(1) = 1e = e\) (since \(A\) is an \(R\)-module). Thus, \(\alpha\) is an identity preserving ring homomorphism. If \(r \in R, a \in A\),

\[ \alpha(r)a = (re)a = r(ea) = ra = r(ae) = a(re) = a\alpha(r). \]

\(b)\) Exercise. \(\square\)
2.3 Remark. The homomorphism \( \alpha : R \to A \) in the proposition above us called the \textit{structural homomorphism} of the associative \( R \)-algebra with identity \( A \). Naturally, a ring \( A \) can have several different \( R \)-algebra structures (corresponding to different structural homomorphisms).

2.4 Examples. a) The ring \( M_n(R) \) of all square matrices of type \( n \times n \) with entries in \( R \) is an associative \( R \)-algebra with identity (noncommutative if \( n \geq 2 \)). The structural homomorphism takes \( r \in R \) to the matrix having \( r \) on the diagonal and 0 elsewhere.

b) The polynomial ring \( R[X] \) is a commutative \( R \)-algebra. If \( K \subseteq L \) is a field extension, \( L \) is a \( K \)-algebra. Which are the structural homomorphisms (equivalently, which is the module structure) for these examples?

2.5 Definition. Let \( A \) and \( B \) be two \( R \)-algebras. A ring homomorphism \( \varphi : A \to B \) that is simultaneously an \( R \)-module homomorphism is called an \( R \)-\textit{algebra homomorphism}.

If \( A \) and \( B \) are associative and have identities, and \( \alpha \), respectively \( \beta \), are the structural homomorphisms, then a ring homomorphism \( \varphi : A \to B \) is an \( R \)-algebra homomorphism if and only if \( \varphi \circ \alpha = \beta \).

In what follows, we consider only \textit{associative algebras with identity}.

A subset \( C \) of the \( R \)-algebra \( A \) is called an \( R \)-\textit{subalgebra} of \( A \) if \( C \) is a subring in \( A \) and also a submodule in \( R \): \( \forall r \in R, \forall a \in C \Rightarrow ra \in C \). It is immediate that the intersection of a family of subalgebras of \( A \) is still a subalgebra of \( A \) (for the proof, see the corresponding property for rings or for modules). This allows to define, given a subset \( S \) of \( A \), the \textit{subalgebra generated by} \( S \) as the intersection of all subalgebras of \( A \) that include \( S \). For \textit{commutative} \( R \)-algebras, the subalgebra generated by \( S \) is denoted by \( R[S] \) and it is the set of all
polynomial expressions in the elements of $S$, with coefficients in $R$ (cf. prop. IV.1.14).

If $\varphi : A \rightarrow B$ is an $R$-algebra homomorphism, then $\varphi(A)$ is a subalgebra of $B$. An ideal $I$ of the ring $A$ is also called an ideal of the $R$-algebra $A$. If $I$ is an ideal of the $R$-algebra $A$, then the factor ring $A/I$ is canonically an $R$-algebra, its structural homomorphism being $\pi \circ \alpha$, where $\pi : A \rightarrow A/I$ is the canonical surjection. This algebra is called the factor algebra of $A$ relative to the ideal $I$.

We fix a commutative ring with identity $R$ and a monoid $(G, \cdot)$. We shall now describe the construction of the monoid algebra $R[G]$. Recall that $(G, \cdot)$ is a monoid (or a semigroup with identity) if the operation "\cdot" is associative and has an identity element $e$.

In particular, we obtain the construction of polynomial algebras (in a finite or infinite number of indeterminates). The idea is the following: we define on $R^G$ (the free $R$-module on the set $G$) a multiplication law that is associative and distributive, and that coincides with the multiplication in $G$ for the elements of $G$. More precisely, any element in $R^G$ is written uniquely as a finite sum
\[ \sum_{g \in G} a_g g, \]
where $a_g \in R$, for any $g \in G$ and $\text{supp}(a_g)_G$ is finite.

The product between $g$ and $h \in G$ (seen as elements in the basis of $R^G$) is $gh$ (element in the basis of $R^G$); this product extends by linearity to any two elements of $R^G$:
\[ \left( \sum_{g \in G} a_g g \right) \cdot \left( \sum_{h \in G} b_h h \right) = \sum_{(g,h) \in G \times G} a_g b_h gh \]

The rigorous construction is described in the sequel.

\[ ^{80} \text{Called group algebra if } G \text{ is a group.} \]
The *support* of a function $\varphi : G \to R$ is the set $\text{supp}(\varphi) := \{g \in G \mid \varphi(g) \neq 0\}$. Define

$$R[G] := \{\varphi : G \to R \mid \text{supp}(\varphi) \text{ is finite}\}.$$ 

A function in $R[G]$ is called a function of *finite support*. We define on $R[G]$ the following operations:

$$\forall \varphi, \psi \in R[G], \quad \forall g \in G,$$

$$(\varphi + \psi)(g) := \varphi(g) + \psi(g)$$

$$(\varphi \cdot \psi)(g) := \sum_{(u,v) \in G \times G} \varphi(u)\psi(v).$$

The first equality clearly defines a function $\varphi + \psi : G \to R$. It is not as obvious that $\varphi \cdot \psi$ is correctly defined: we must show that the sum defining $\varphi \cdot \psi$ has a finite number of nonzero terms. Indeed, the set of all $(u,v) \in G \times G$ such that $\varphi(u)\psi(v) \neq 0$ is included in $\text{supp}(\varphi) \times \text{supp}(\psi)$, which is finite.

We must also show that $\varphi + \psi$ and $\varphi \cdot \psi$ have finite support. We have:

$$\text{supp}(\varphi + \psi) \subseteq \text{supp}(\varphi) \cup \text{supp}(\psi),$$

a finite set.

For $\varphi \cdot \psi$: if $g \in G \setminus \{uv \mid u \in \text{supp}(\varphi) \text{ and } v \in \text{supp}(\psi)\}$, then $(\varphi \cdot \psi)(g) = 0$, since all terms in the sum in the definition are zero. So, $\text{supp}(\varphi \psi)$ is included in $\{uv \mid u \in \text{supp}(\varphi) \text{ and } v \in \text{supp}(\psi)\}$, which is finite.

Therefore, "+" and "·" are correctly defined and are composition laws on $R[G]$. We define an external operation ",": $r \cdot \varphi : R \times R[G] \to R[G]$,

$$(r \cdot \varphi)(g) := r\varphi(g), \quad \forall r \in R, \forall \varphi \in R[G], \forall g \in G.$$

Endowed with these operations, $R[G]$ is an $R$-module, namely the *free $R$-module of basis* $G$ (if we disregard the multiplication in $R[G]$).

2.1 Remark. The construction we described generalizes (and is inspired from) the usual construction of the polynomial ring in one indeterminate over $R$, $R[X]$. The reader is encouraged to verify that if -
(G, ·) is (\(\mathbb{N}, +\)), the additive monoid of the natural numbers, then \(R[G]\) is exactly the ring \(R[X]\).

### 2.2 Proposition. \((R[G], +, \cdot)\) is an associative ring with identity.

**Proof.** We check the associativity of multiplication. Let \(\varphi, \psi, \eta \in R[G]\) and let \(g \in G\).

\[
((\varphi\psi)\eta)(g) = \sum_{(u,v)\in G^2 \atop uv=g} (\varphi\psi)(u)\eta(v) = \sum_{(u,v)\in G^2 \atop uv=g} \varphi(s)\psi(t)\eta(v) = \sum_{(s,t,v)\in G^3 \atop stv=g} \varphi(s)\psi(t)\eta(v).
\]

Computing \((\varphi(\psi\eta))(g)\), we obtain the same thing, so \((\varphi\psi)\eta = \varphi(\psi\eta)\).

The existence of neutral elements for addition and multiplication is proven below. The other ring axioms are left to the reader. \(\square\)

It is necessary to show that this construction satisfies the demands stated at the beginning of this section and. To this end, we define the following elements in \(R[G]\):

- \(\forall g \in G\), define \(\eta_g : G \to R\) by \(\eta_g(h) = \begin{cases} 0, & \text{if } h \neq g \\ 1, & \text{if } h = g \end{cases}, \forall h \in G;\)

- \(\forall r \in R\), define \(\psi_r : G \to R\) by \(\psi_r(h) = \begin{cases} 0, & \text{if } h \neq e \\ r, & \text{if } h = e \end{cases}, \forall h \in G;\)

It is clear that \(\eta_g, \psi_r \in R[G], \forall g \in G, \forall r \in R\). We also have:

### 2.3 Proposition. a) The function \(i : R \to R[G]\), given by \(i(r) = \psi_r\), \(\forall r \in R\), is an injective ring homomorphism. Furthermore, \(\text{Im } i\) is included in the center of \(R[G]\) (so, \(R[G]\) is an \(R\)-algebra of structural homomorphism \(i\)). That entitles us to write \(r\) instead of \(\psi_r\) (identifying \(r \in R\) with its image \(\psi_r \in R[G]\)).
b) The function \( j : G \to (R[G], \cdot), j(g) = \eta_g, \ \forall g \in G \), is an injective monoid homomorphism. We write \( g \) instead of \( \eta_g \) (identifying \( g \in G \) with its image \( \eta_g \in R[G] \)).

c) For any \( g, h \in G \) and any \( r \in R \), \( (\psi_r \cdot \eta_g)(h) = \begin{cases} 0, & \text{if } h \neq g \\ r, & \text{if } h = g \end{cases} \).

d) Any element \( \varphi \in R[G] \) is written as a finite sum:

\[
\varphi = \sum_{g \in \supp(\varphi)} \psi_{\varphi(g)} \eta_g = \sum_{g \in \supp(\varphi)} a_g g.
\]

In the second sum, \( \varphi(g) \) is denoted \( a_g \), \( \eta_g \) is identified with \( g \) and \( \psi_{\varphi(g)} \) is identified with \( a_g = \varphi(g) \), for any \( g \in G \) (thus \( (a_g)_{g \in G} \) has finite support).

The writing of \( \varphi \) is unique: if \( \sum_{g \in G} a_g g = \sum_{g \in G} b_g g \), where \( (a_g)_{g \in G} \) and \( (b_g)_{g \in G} \) have finite support, then \( a_g = b_g, \ \forall g \in G \).

e) The zero of the ring \( R[G] \) (the neutral element for addition) is \( \psi_0 \) (written as a sum of the type \( \sum_{g \in G} a_g g \) as the sum with one term \( 0e \)).

The identity of the ring \( R[G] \) (the neutral element for multiplication) is \( \eta_e = 1e \).

**Proof.**

a) It is evident that \( \psi_{r+s} = \psi_r + \psi_s, \ \forall r, s \in R \). Computing \( \psi_r \cdot \psi_s \), we obtain \( \psi_r \cdot \psi_s(g) = \sum_{uv=g} \psi_r(u)\psi_s(v) \). If \( g \neq e \), then, for any couple \( (u, v) \in G \times G \) such that \( uv = g \), we have \( u \neq e \) or \( v \neq e \), so \( \psi_r(u)\psi_s(v) = 0 \). Therefore, if \( g \neq e \), then \( (\psi_r \cdot \psi_s)(g) = 0 \). Also, \( (\psi_r \cdot \psi_s)(e) = \psi_r(e) \cdot \psi_s(e) = rs \). In conclusion, \( \psi_r \cdot \psi_s = \psi_{rs} \). The injectivity is clear.

b) The injectivity is easy. We prove that \( \eta_g \eta_h = \eta_{gh}, \ \forall g, h \in G \). For any \( x \in G, x \neq gh \), \( (\eta_g \eta_h)(x) = \sum_{uv=x} \eta_g(u)\eta_h(v) = 0 \), since \( uv = x \neq gh \) implies \( u \neq g \) or \( v \neq h \). Also, \( (\eta_g \eta_h)(gh) = 1 \).
c) Exercise.

d) For any $h \in G$, we have

\[
\left( \sum_{g \in \text{supp} \varphi} a_g g \right)(h) = \sum_{g \in \text{supp}(\varphi)} (\psi_{\varphi(g)} \eta_g)(h) = \begin{cases} 0, & \text{if } h \notin \text{supp}(\varphi) \\ \varphi(h), & \text{if } h \in \text{supp}(\varphi) \end{cases} = \varphi(h).
\]

We used in the last equality that \( (\psi_{\varphi(g)} \eta_g)(h) = \begin{cases} 0, & \text{if } h \neq g \\ \varphi(g), & \text{if } h = g \end{cases} \), as seen at c).

The uniqueness ensues from \( \left( \sum_{g \in G} a_g g \right)(h) = a_h, \forall h \in G \).

e) We show that \( \eta_e \) is the identity of the ring \( R[G] \). For any \( g \in G \), \( \eta_g \eta_e = \eta_{ge} = \eta_g = \eta_e \eta_g \), by c). The general case is proven using d) and the distributivity.

\[ \Box \]

2.4 Remarks. a) By construction, \( R[G] \) is isomorphic to the free \( R \)-module of basis \( G \). The elements of \( R[G] \) can be seen as “formal” finite sums of the form \( \sum_{g \in G} a_g g \), where \( (a_g)_{g \in G} \) is a family of elements of \( R \) having finite support. We can identify \( a \in R \) with the sum with one term \( a \cdot e \); also, we can identify \( g \in G \) with \( 1 \cdot g \). The addition obeys the rule \( \sum_{g \in G} a_g g + \sum_{g \in G} b_g g = \sum_{g \in G} (a_g + b_g)g \); the multiplication is (left and right) distributive relative to addition and obeys the rule \( (1 \cdot g)(1 \cdot h) = 1 \cdot (gh) \). Therefore:

\[
\left( \sum_{g \in G} a_g g \right) \cdot \left( \sum_{g \in G} b_g g \right) = \sum_{g \in G} \left( \sum_{uv=g} (a_u b_v) \right) g .
\]

Thus, \( R[G] \) meets the conditions stated in the introduction. Any element of \( R[G] \) is written uniquely \( \sum_{g \in G} a_g g \), the sums being always finite. In particular, \( \sum_{g \in G} a_g g = 0 \iff a_g = 0, \forall g \in G \).
b) If $G$ is a commutative monoid, then $R[G]$ is also a commutative ring. If $G$ is not commutative, then $R[G]$ is not commutative, as part $b$ of the proposition shows.

The classical polynomial algebras.

I. For $(G, \cdot) = (\mathbb{N}, +)$, one obtains the usual construction of the polynomial algebra in one indeterminate with coefficients in $R$. Indeed, an arbitrary element in $R[\mathbb{N}]$ is a function $\varphi : \mathbb{N} \to R$ with finite support (a sequence of elements of $R$ with finite support). Setting $\varphi(i) = a_i$, $\forall i \in \mathbb{N}$, the general form of an element $f$ in $R[\mathbb{N}]$ is $f = \sum_{i \in \mathbb{N}} a_i \eta_i$. Since $\eta_i \eta_j = \eta_{i+j}$, for any $i, j \in \mathbb{N}$, $\eta_i = (\eta_1)^i$, $\forall i \in \mathbb{N}$.

Denoting $\eta_1$ by $X$, one obtains the usual form:

$$f = \sum_{i \in \mathbb{N}} a_i X^i = a_0 + a_1 X + \ldots + a_n X^n$$

where $n = \max \{i \in \mathbb{N} \mid a_i \neq 0\}$ is the degree of $f$.

II. Consider the commutative monoid $(\mathbb{N}^n, +)$ (for a fixed $n \in \mathbb{N}^*$), where the addition is defined component-wise:

$$(i_1, \ldots, i_n) + (j_1, \ldots, j_n) := (i_1 + j_1, \ldots, i_n + j_n), \forall (i_1, \ldots, i_n),$$

$$(j_1, \ldots, j_n) \in \mathbb{N}^n.$$

The $R$-algebra $R[\mathbb{N}^n]$ is called the polynomial algebra in $n$ indeterminates\(^{81}\) with coefficients in $R$. An element of $R[\mathbb{N}^n]$ is called a polynomial (in $n$ indeterminates).

We make the connection with the usual form of a polynomial. For each $i \in \{1, \ldots, n\}$, let

$$e_i := (0, \ldots, 1, \ldots, 0) \in \mathbb{N}^n \ (1 \text{ on the } i\text{-th place, } 0 \text{ elsewhere}).$$

\(^{81}\) Often encountered terminologies are “unknown” or “variable” instead of “indeterminate”.
Any element in $\mathbb{N}^n$ is written uniquely - up to a order of the terms - as a sum of $e_i$'s (in other words, $\{e_1, \ldots, e_n\}$ generate the monoid $\mathbb{N}^n$). The element $\eta_{e_i} \in R[\mathbb{N}^n]$ is denoted by $X_i$ and is called an indeterminate. A product of indeterminates (of the form $X_1^{i_1} \ldots X_n^{i_n}$) is called a term. Any polynomial $g \in R[\mathbb{N}^n]$ is then written uniquely as a finite sum:

$$g = \sum_{(i_1, \ldots, i_n) \in \mathbb{N}^n} a_{i_1 \ldots i_n} X_1^{i_1} \ldots X_n^{i_n},$$

where $\left(a_{i_1 \ldots i_n}\right)_{(i_1, \ldots, i_n) \in \mathbb{N}^n}$ is a family of elements in $R$, having finite support. Thus, $g$ is a linear combination of terms, with coefficients in $R$. The polynomial $g$ is thus a sum of products between a nonzero element of $R$ and a term, of the form $a_{i_1 \ldots i_n} X_1^{i_1} \ldots X_n^{i_n}$. If $g$ is like above and $a_{i_1 \ldots i_n} \neq 0$, then $a_{i_1 \ldots i_n} X_1^{i_1} \ldots X_n^{i_n}$ is called a monomial of $g$.

$R[\mathbb{N}^n]$ is denoted usually by $R[X_1, \ldots, X_n]$.

III. The previous construction can be generalized to the polynomial algebra in $S$ indeterminates, where $S$ is an arbitrary nonempty set. To this end, we replace the monoid $(\mathbb{N}^n, +)$ with $(\mathbb{N}^{(S)}, +)$, where $\mathbb{N}^{(S)} := \{f : S \rightarrow \mathbb{N} \mid \text{the support of } f \text{ is finite}\}$. Continuing the analogy with $\mathbb{N}^n$, we see the elements of $\mathbb{N}^{(S)}$ as “multi-indices” and use notations like $i, j, \ldots$. The addition in the monoid $\mathbb{N}^{(S)}$ is defined naturally as follows: for any $i, j \in \mathbb{N}^{(S)}$, $(i + j)(s) = i(s) + j(s), \forall s \in S$. The axioms of a commutative monoid are readily checked. The $R$-algebra $R[\mathbb{N}^{(S)}]$ is called the polynomial algebra in $S$ indeterminates with coefficients in $R$. For any $s \in S$, define $e_s \in \mathbb{N}^{(S)}$, $e_s(t) = \begin{cases} 0, & \text{if } s = t \\ 1, & \text{if } s \neq t \end{cases}$, $\forall t \in S$ and let $X_s$ be the element $\eta_{e_s} \in R[\mathbb{N}^{(S)}]$. 
Any element \( i \in \mathbb{N}^{(S)} \) is written uniquely as \( i = \sum_{s \in S} m_s e_s \), where \((m_s)_{s \in S}\) is a family of natural numbers indexed by \( S \), having finite support. Therefore, \( \eta_i = \prod_{s \in \text{supp}(i)} X_s^{m_s} \). An arbitrary polynomial \( f \in R[\mathbb{N}^{(S)}] \) has a unique writing as \( f = \sum_{i \in F} a_i \eta_i \), where \( F \) is a finite subset of \( \mathbb{N}^{(S)} \); if \( \bigcup_{i \in F} \text{supp}(i) = \{s_1, \ldots, s_n\} \) (a finite subset of \( S \)), then

\[
f = \sum_{(m_1, \ldots, m_n) \in \mathbb{N}^n} a_{m_1 \ldots m_n} X_{s_1}^{m_1} \ldots X_{s_n}^{m_n},
\]

where the sum is finite (the family of elements of \( R \) \((a_{m_1 \ldots m_n})_{(m_1, \ldots, m_n) \in \mathbb{N}^n} \) has finite support).

In other words, **any polynomial in \( S \) indeterminates is a polynomial in a finite set of indeterminates in \( S \)**. The \( R \)-algebra \( R[\mathbb{N}^{(S)}] \) is usually denoted as \( R[\langle X_s \rangle_{s \in S}] \) or \( R[X_s]_{s \in S} \) or \( R[X; S] \).

**2.5 Theorem.** (The universality property of the monoid algebra) Let \((G, \cdot)\) be a monoid and let \( i : R \to R[G], j : G \to R[G] \) be the canonical mappings defined at 2.3. The triple \((R[G], i, j)\) has the following universality property: for any \( R \)-algebra \( T \) having the structural homomorphism \( \alpha : R \to T \) and any homomorphism of monoids \( \beta : G \to (T, \cdot), \) there exists a unique homomorphism of \( R \)-algebras \( \varphi : R[G] \to T \) such that \( \varphi \circ i = \alpha \) and \( \varphi \circ j = \beta \):

\[
\begin{array}{ccc}
G & \xrightarrow{j} & R[G] \\
\downarrow{\beta} & & \downarrow{\varphi} \\
T & & \\
\end{array}
\]

---

82 For any \( m \in \mathbb{N} \) and \( i \in R[\mathbb{N}^{(S)}] \), \( mi \) is defined as \((mi)(s) := m \cdot i(s), \forall s \in S\).
Proof. Suppose \( \varphi \) is as stated. Thus, \( \varphi(r) = \alpha(r), \ \forall r \in R \) and \( \varphi(g) = \beta(g), \ \forall g \in G \). If \( \sum_{g \in G} a_g g \in R[G] \), then

\[
\varphi\left(\sum_{g \in G} a_g g\right) = \sum_{g \in G} \varphi(a_g) \varphi(g) = \sum_{g \in G} \alpha(a_g) \beta(g),
\]

which shows that \( \varphi \) is uniquely determined by \( \alpha \) and \( \beta \). A direct check shows that \( \varphi \) given by the above is a ring homomorphism and satisfies the stated requirements.

The universality property of the monoid algebra determines this algebra up to an unique isomorphism: if \( (A, \gamma, \delta) \) (where \( A \) is an \( R \)-algebra, \( \gamma : R \to A \) is its structural homomorphism and \( \delta : G \to (A, \cdot) \) is a homomorphism of monoids) satisfies the same universality property as \( (R[G], i, j) \), then there exists a unique \( R \)-algebra isomorphism \( \varphi : R[G] \to A \) such that \( \varphi i = \gamma \) and \( \varphi j = \delta \) (cf. 5.7).

For the classical polynomial algebras, using the remarkable fact that the monoid \( (\mathbb{N}, +) \) is generated by one element (respectively \( (\mathbb{N}^n, +) \) is generated by \( n \) elements), this theorem reads:

2.6 Theorem. Let \( A \) be an \( R \)-algebra.

\( a) \) (The universality property of the polynomial algebra \( R[X] \)) For any \( a \in A \) there exists a unique \( R \)-algebra homomorphism \( \text{ev}_a : R[X] \to A \) such that \( \text{ev}_a(X) = a \).

\( b) \) (The universality property of the polynomial algebra \( R[X_1, \ldots, X_n] \)) Let \( n \in \mathbb{N}^* \). For any \( n \)-uple \( a = (a_1, \ldots, a_n) \in A^n \) there exists a unique \( R \)-algebra homomorphism \( \text{ev}_a : R[X_1, \ldots, X_n] \to A \) such that \( \text{ev}_a(X_i) = a_i, \ \forall i \in \{1, \ldots, n\} \).

\( c) \) (The universality property of the polynomial algebra \( R[X; S] \)) Let \( S \) be a nonempty set. For any mapping \( \gamma : S \to A \) there exists a unique \( R \)-algebra homomorphism \( \text{ev}_\gamma : R[X; S] \to A \) such that \( \text{ev}_\gamma(X_s) = \gamma(s), \ \forall s \in S \).
2. Algebras. Polynomial and monoid algebras

**Proof.** Of course, (a) follows from (b), which is a particular case of (c). Nevertheless, we sketch a proof for every case.

(a) For any \( f = \sum_{i=0}^{n} b_i X^i \in R[X] \), let \( ev_a(f) = \sum_{i=0}^{n} b_i a^i \) (usually called the value of \( f \) at \( a \) and denoted \( f(a) \)). The fact that \( ev_a \) is a homomorphism amounts to that, \( \forall f, g \in R[X] \),

\[
(f + g)(a) = f(a) + g(a),
\]

\[
(f \cdot g)(a) = f(a) \cdot g(a),
\]

A standard check proves these familiar properties and the rest of the claims.

(b) For any \( f = \sum_{(i_1, \ldots, i_n) \in \mathbb{N}^n} b_{i_1 \cdots i_n} X_1^{i_1} \cdots X_n^{i_n} \in R[X_1, \ldots, X_n] \), and any \( a = (a_1, \ldots, a_n) \in A^n \), let

\[
ev_a(f) = f(a_1, \ldots, a_n) = \sum_{(i_1, \ldots, i_n) \in \mathbb{N}^n} b_{i_1 \cdots i_n} a_1^{i_1} \cdots a_n^{i_n},
\]

the value of \( f \) at \( a = (a_1, \ldots, a_n) \).

(c) The mapping \( \gamma \) induces a homomorphism of monoids

\[
\beta : \mathbb{N}^{(S)} \to (A, \cdot), \beta(i) = \prod_{s \in \text{supp}(i)} \gamma(s)^{i(s)}, \forall i \in \mathbb{N}^{(S)}
\]

Applying the universality property of the monoid algebra \( R[\mathbb{N}^{(S)}] = R[X; S] \), there exists a unique \( R \)-algebra homomorphism \( ev_\gamma : R[X; S] \to A \) such that \( \nu \circ j = \beta \), where \( j : \mathbb{N}^{(S)} \to R[X; S] \) is the canonical mapping; in our case, \( j(e_s) = X_s, \forall s \in S \). Thus \( \nu_\gamma(X_s) = \beta(e_s) = \gamma(s), \forall s \in S \).

Let us prove the uniqueness of \( ev_\alpha \). If \( \nu : R[\mathbb{N}^{(S)}] \to A \) is an \( R \)-algebra homomorphism with \( \nu(X_s) = \gamma(s) \), then \( \nu \circ j = \beta \), where \( \beta \) is the homomorphism defined above. The uniqueness part of the universality property implies \( \nu = ev_\alpha \). \( \square \)
The homomorphism \( ev_a \) (respectively \( ev_a \)) above is called the evaluation homomorphism.

The previous theorem formalizes and gives a precise meaning to the procedure of assigning the values \( a_1, \ldots, a_n \) to the indeterminates \( X_1, \ldots, X_n \).

A useful property of the \( R \)-algebras \( R[X_1, \ldots, X_n] \), sometimes used to define them by recurrence on \( n \), is the following:

2.7 Theorem. Let \( n \geq 1 \). Then there exists a canonical isomorphism of \( R \)-algebras:

\[
R[X_1, \ldots, X_n] \cong R[X_1, \ldots, X_{n-1}][X_n].
\]

Proof. Use 2.6.b): there exists a unique \( R \)-algebra homomorphism

\[
\varphi: R[X_1, \ldots, X_n] \to R[X_1, \ldots, X_{n-1}][X_n], \text{ with } \varphi(X_i) = X_i, \ 1 \leq i \leq n.
\]

Let \( A := R[X_1, \ldots, X_{n-1}] \).

Conversely, 2.6.b) applied to \( R[X_1, \ldots, X_{n-1}] \), yields a unique \( R \)-algebra homomorphism \( \alpha: R[X_1, \ldots, X_{n-1}] \to R[X_1, \ldots, X_n] \), with \( \alpha(X_i) = X_i, \ 1 \leq i \leq n - 1 \). Thus, \( R[X_1, \ldots, X_n] \) is an \( A \)-algebra with \( \alpha \) its structural homomorphism. The universality property of the \( A \)-algebra \( A[X_n] \) shows that there exists a unique \( A \)-algebra homomorphism \( \beta: A[X_n] \to R[X_1, \ldots, X_n] \), such that \( \beta(X_n) = X_n \). Of course, \( \beta \) is also an \( R \)-algebra homomorphism.

We have \( \beta \varphi = \text{id} \). Indeed, \( \beta \varphi: R[X_1, \ldots, X_n] \to R[X_1, \ldots, X_n] \) is an \( R \)-algebra homomorphism such that \( \beta \varphi(X_i) = X_i, \ 1 \leq i \leq n \), and \( \text{id}: R[X_1, \ldots, X_n] \to R[X_1, \ldots, X_n] \) has the same properties. The uniqueness part in 2.6.b) implies that \( \beta \varphi = \text{id} \). Similarly, \( \varphi \beta = \text{id} \), so \( \varphi \) is an isomorphism.

In what follows we discuss the important notion of degree of a polynomial.

If \( aX^n \) is a monomial in \( R[X] \) (with \( a \neq 0 \)), \( n \) is called the degree of \( aX^n \).

For an arbitrary polynomial \( g \in R[X] \),
\[ g = a_0 + a_1X + \ldots + a_nX^n, \text{ where } a_n \neq 0, \]

the natural number \( n \) is called the degree of \( g \), denoted \( \deg g \). Thus, the degree of \( g \) is the greatest degree of the monomials of \( g \). We define by convention \( \deg 0 = -\infty \). Sometimes \( \deg 0 \) is not defined.

The elements \( a_0, \ldots, a_n \in R \) are called the coefficients of the polynomial \( g \); among these, \( a_n \) is called the leading coefficient of \( g \), \( a_0 \) is called the constant term.

If the leading coefficient \( a_n \) is 1, then the polynomial \( g \) is called monic.

If \( R \) is a domain, then the degree is additive: \( \forall g, h \in R[X_1,\ldots,X_n], \)
\[ \deg (gh) = \deg g + \deg h. \]

Also:
\[ \deg (g + h) \leq \max(\deg g, \deg h). \]

If \( aX_{i_1}^{i_1} \ldots X_{i_n}^{i_n} \) is a monomial in \( R[X_1,\ldots,X_n] \) (where \( a \neq 0 \), and \( 1 \leq k \leq n \), then its degree in \( X_k \) is \( \deg \left( aX_{i_1}^{i_1} \ldots X_{i_n}^{i_n}, X_k \right) := i_k \) (the exponent of \( X_k \) in the monomial), also denoted \( \deg_{X_k} \left( aX_{i_1}^{i_1} \ldots X_{i_n}^{i_n} \right) \).

For any \( g \in R[X_1,\ldots,X_n] \), \( \deg \left( g, X_k \right) \) is the greatest degree in \( X_k \) of the monomials of \( g \). If \( R \) is a domain, then the degree in \( X_k \) satisfies the same relations as above: \( \forall g, h \in R[X_1,\ldots,X_n], \)
\[ \deg \left( gh, X_k \right) = \deg \left( g, X_k \right) + \deg \left( h, X_k \right) \]
\[ \deg \left( g + h, X_k \right) \leq \max(\deg \left( g, X_k \right), \deg \left( h, X_k \right)). \]

The total degree of the monomial \( aX_{i_1}^{i_1} \ldots X_{i_n}^{i_n} \) is \( i_1 + \ldots + i_n \); the total degree of an arbitrary polynomial \( g \in R[X_1,\ldots,X_n] \) is the largest total degree of its monomials. Usually, the “degree” of a polynomial in several indeterminates is its total degree, unless otherwise specified. A polynomial whose monomials have all the same degree is called a homogeneous polynomial. The total degree satisfies relations similar to the above, if \( R \) is a domain.
3. Symmetric polynomials

Let $R$ be a fixed commutative ring with identity. Let $n \in \mathbb{N}^*$ and let $\sigma \in S_n$, the symmetric group of all permutations of $n$ objects. Then there exists a unique homomorphism of $R$-algebras $\varphi_\sigma : R[X_1, \ldots, X_n] \to R[X_1, \ldots, X_n]$ such that $\varphi_\sigma(X_i) = X_{\sigma(i)}$, $\forall i = 1, \ldots, n$ (see 2.6, the universality property of the polynomial $R$-algebra $R[X_1, \ldots, X_n]$). If $g \in R[X_1, \ldots, X_n]$, then $\varphi_\sigma(g) = g(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$.

If $R$ is a domain and $K$ is its field of fractions, consider $K(X_1, \ldots, X_n)$ (the field of fractions of the domain $R[X_1, \ldots, X_n]$, called the field of rational fractions in the indeterminates $X_1, \ldots, X_n$ with coefficients in $K$). Then $\varphi_\sigma$ extends to a unique field homomorphism (denoted also $\varphi_\sigma$) $\varphi_\sigma : K(X_1, \ldots, X_n) \to K(X_1, \ldots, X_n)$. For any $g, h \in R[X_1, \ldots, X_n], h \neq 0$, $\varphi_\sigma(g/h) = \varphi_\sigma(g)/\varphi_\sigma(h)$.

3.1 Definition. Let $g \in R[X_1, \ldots, X_n]$. We call $g$ a symmetric polynomial in $R[X_1, \ldots, X_n]$ if, for any $\sigma \in S_n$, we have $\varphi_\sigma(g) = g$.

If $R$ is a domain and $K$ is its field of fractions, a rational fraction $g/h \in K(X_1, \ldots, X_n)$ is called symmetric if, for any $\sigma \in S_n$, $\varphi_\sigma(g/h) = g/h$.

3.2 Example. In $R[X_1, X_2, X_3]$, the following polynomials are symmetric:

$$X_1 + X_2 + X_3, \ X_1 X_2 X_3,$$

$$X_1^2 X_2 + X_1^2 X_3 + X_2^2 X_1 + X_2^2 X_3 + X_3^2 X_1 + X_3^2 X_2.$$

The polynomial $X_1 + X_2$ is not symmetric in $R[X_1, X_2, X_3]$ (but it is symmetric in $R[X_1, X_2]$).

3.3 Remarks. a) Let $S = \{g \in R[X_1, \ldots, X_n] \mid \varphi_\sigma(g) = g, \forall \sigma \in S_n\}$ be the set of symmetric polynomials. Then $S$ is an $R$-subalgebra of $R[X_1, \ldots, X_n]$. For instance, if $g, h \in S$, then
3. Symmetric polynomials

\[ \varphi_{\sigma}(g + h) = \varphi_{\sigma}(g) + \varphi_{\sigma}(h) = g + h, \forall \sigma \in S_n. \]

The other conditions are checked similarly. It is easy to see that the symmetric rational fractions form a subfield in \( K(X_1, \ldots, X_n) \).

b) If \( aX_{i_1}^{r_{i_1}} \ldots X_{i_n}^{r_{i_n}} \) is a monomial of the symmetric polynomial \( g \), then \( aX_{\sigma(1)}^{r_{\sigma(1)}} \ldots X_{\sigma(n)}^{r_{\sigma(n)}} \) is also a monomial of \( g \), for any \( \sigma \in S_n \).

3.4 Definition. Let \( n \in \mathbb{N}^* \) and let \( 0 \leq k \leq n \). The polynomial

\[ s_k := \sum \{ \prod_{i \in I} X_i \mid I \subseteq \{1, \ldots, n\}, |I| = k \} \]

is called the fundamental (or elementary) symmetric polynomial of degree \( k \) in \( R[X_1, \ldots, X_n] \).

In other words, \( s_k \) is the sum of all products of \( k \) distinct indeterminates chosen among \( \{X_1, \ldots, X_n\} \); thus \( s_k \) has \( \binom{n}{k} \) monomials. By convention, \( s_0 = 1 \) and \( s_k = 0 \) if \( k > n \). The polynomial \( s_k \) is homogeneous of degree \( k \) (indeed, all its monomials have degree \( k \)). Since \( s_k \) obviously depends on the number of indeterminates, the notation \( s_k(X_1, \ldots, X_n) \) is sometimes used to avoid confusions. For example, the fundamental symmetric polynomials in 4 indeterminates are:

\[
\begin{align*}
  s_0 &= 1 \\
  s_1 &= X_1 + X_2 + X_3 + X_4 \\
  s_2 &= X_1X_2 + X_1X_3 + X_1X_4 + X_2X_3 + X_2X_4 + X_3X_4 \\
  s_3 &= X_1X_2X_3 + X_1X_2X_4 + X_1X_3X_4 + X_2X_3X_4 \\
  s_4 &= X_1X_2X_3X_4
\end{align*}
\]

The fundamental symmetric polynomials appear in the relations between the coefficients of a polynomial and its roots (Viète's relations).

3.5 Theorem. a) Let \( n \in \mathbb{N}^* \) and let \( s_k = s_k(X_1, \ldots, X_n) \). In \( R[X_1, \ldots, X_n][X] \) the following relation holds:

\[
(X - X_1) \cdots (X - X_n) = X^n - s_1X^{n-1} + s_2X^{n-2} - \ldots + (-1)^n s_n.
\]
b) If $R$ is a subring of the domain $S$ and $g = a_0 + a_1X + \ldots + a_nX^n \in R[X]$ has the roots $x_1, \ldots, x_n \in S$, then $a_ns_k(x_1, \ldots, x_n) = (-1)^k a_{n-k}$.

**Proof.** a) Induction on $n$ (exercise).

b) There exists a unique $R$-algebra homomorphism $\varphi : R[X_1,\ldots,X_n][X] \to S[X]$ such that $\varphi(X_i) = x_i$ and $\varphi(X) = X$. We have, by a):

$$\varphi(a_n(X-X_1)\ldots(X-X_n)) = a_n(X-x_1)\ldots(X-x_n) = a_nX^n - s_1X^{n-1} + s_2X^{n-2} - \ldots + (-1)^ns_n.$$ 

But $a_n(X-x_1)\ldots(X-x_n) = g$ (in the field of fractions $K$ of $S$, these polynomials have the same roots and the same leading coefficient). The relations follow by identifying the coefficients. \hfill $\square$

### 3.6 Theorem

(The fundamental theorem of symmetric polynomials) Let $g$ be a symmetric polynomial in $R[X_1,\ldots,X_n]$. Then $g$ is a polynomial of the fundamental symmetric polynomials: there exists a unique polynomial $h \in R[X_1,\ldots,X_n]$ such that $g = h(s_1, \ldots, s_n)$.

In other words, denoting by $S$ the $R$-subalgebra of symmetric polynomials in $R[X_1,\ldots,X_n]$, the unique $R$-algebra homomorphism $\psi : R[X_1,\ldots,X_n] \to S$ with $\psi(X_i) = s_i$ (for $1 \leq i \leq n$) is an isomorphism.

**Proof.** Let $T := \{X_1^{i_1}\ldots X_n^{i_n} | (i_1,\ldots,i_n) \in \mathbb{N}^n\}$ be the set of all terms in $R[X_1,\ldots,X_n]$. Define a total order relation on $T$ (the lexicographic order) by: $X_1^{i_1}\ldots X_n^{i_n} \leq X_1^{k_1}\ldots X_n^{k_n} \iff \exists r, 1 \leq r \leq n$, such that $i_t = k_t, \forall t < r$ and $i_r \leq k_r$. For instance, we have $X_1 > X_2 > \ldots > X_n$ and $1 < X_1^3 < X_2X_3^2 < X_1X_1^2$. This order relation is total and it is compatible with term multiplication: $\forall \lambda, \mu, \nu \in T$, $\mu \leq \nu$ implies $\lambda\mu \leq \lambda\nu$ (one can prove that it is the unique total order on $T$, compatible with multiplication, such that $X_1 > X_2 > \ldots > X_n$). Moreover, $T$ is well ordered by the lexicographic order (any nonempty subset of $T$ has a smallest element), as the following lemma shows. Therefore, one can make induction proofs on this ordered set (as is this proof).
The lexicographic order induces a preorder relation\(^{83}\), denoted also "\(\leq\)\), on the set \(\{a\lambda \mid \lambda \in T, a \in R, a \neq 0\}\) of all monomials in \(R[X_1, \ldots , X_n]\), by \(a\lambda \leq b\mu \iff \lambda \leq \mu\). The proof of this is straightforward. If \(p \in R[X_1, \ldots , X_n]\), there exists a unique greatest monomial of \(p\) (with respect to the lexicographic preorder), called the leading monomial of \(p\) and denoted by \(\text{lm}(p)\). The following property holds:

If \(p, q \in R[X_1, \ldots , X_n]\), such that \(\text{lm}(p) = a\lambda, \text{lm}(q) = b\mu\), where \(\lambda, \mu \in T, a, b \in R\) and \(ab \neq 0\), then \(\text{lm}(pq) = \text{lm}(p)\text{lm}(q) = ab\lambda\mu\).

Indeed, any monomial of \(pq\) is a sum of monomials of the form \(r\alpha s\beta\), where \(r\alpha\) is a monomial of \(p\) and \(s\beta\) is a monomial of \(q\). But \(\alpha \leq \lambda\) and \(\beta \leq \mu\), so \(\alpha\beta \leq \lambda\beta \leq \lambda\mu\). Thus, \(ab\lambda\mu = \text{lm}(pq)\).

We proceed to the proof of the theorem. Let \(g\) be a symmetric polynomial and let \(\text{lm}(g) = aX_1^{i_1} \cdots X_n^{i_n}\). Then \(i_1 \geq i_2 \geq \ldots \geq i_n\) (if not, then there exists a \(k\) such that \(i_k < i_{k+1}\), and \(aX_1^{i_1} \cdots X_k^{i_k+1}X_{k+1}^{i_{k+1}} \cdots X_n^{i_n}\) is also a monomial in \(g\), strictly greater than \(\text{lm}(g)\), contradiction). We want to find a polynomial \(p\) of the form \(a_{1}^{i_1} \cdots a_{n}^{i_n}\), such that \(\text{lm}(p) = \text{lm}(g)\). Using the above property,

\[
\text{lm}\left(a_{1}^{i_1} \cdots a_{n}^{i_n}\right) = aX_1^{i_1} (X_1 X_2)^{i_2} \cdots (X_1 \ldots X_n)^{i_n}
\]

This monomial equals \(\text{lm}(g)\) if and only if \(j_1 + \ldots + j_n = i_1, j_2 + \ldots + j_n = i_2, \ldots , j_n = i_n\). Thus \(j_n = i_n, j_k = i_k - i_{k+1}\), for \(1 \leq k < n\). The polynomial

\[
g_1 := g - a_{1}^{i_1} \cdots a_{n}^{i_n}
\]

is symmetric and \(\text{lm}(g_1) < \text{lm}(g)\). If \(\text{lm}(g_1) = 0\), then \(g_1 = 0\) and we are finished. If \(\text{lm}(g_1) \neq 0\), replace \(g\) by \(g_1\) and apply the procedure above. The algorithm terminates in a finite number of steps because any strictly decreasing sequence of terms must be finite, as the next lemma shows. This concludes the existence part of the proof.

Let us prove the uniqueness (or, equivalently, \(\text{Ker}\psi = 0\)). Suppose there exists a nonzero polynomial \(p \in R[X_1, \ldots , X_n]\) such that \(a_{1}^{i_1} \cdots a_{n}^{i_n}\neq 0\). We define a relation that is reflexive and transitive, but not necessarily antisymmetric.

\(^{83}\) A relation that is reflexive and transitive, but not necessarily antisymmetric.
\( \psi(p) = p(s_1, ..., s_n) = 0. \) We claim that there exists a unique nonzero monomial \( \bar{\lambda} \) of \( p \) such that \( \text{lm}(\psi(p)) = \text{lm}(\bar{\lambda}(s_1, ..., s_n)) \). Indeed, if \( \alpha = X_1^{i_1} ... X_n^{i_n} \) and \( \beta = X_1^{j_1} ... X_n^{j_n} \) are distinct terms, then:

\[
\text{lm}(\alpha(s_1, ..., s_n)) = X_1^{i_1+...+i_n} ... X_n^{i_n} \neq X_1^{j_1+...+j_n} ... X_n^{j_n} = \text{lm}(\beta(s_1, ..., s_n)).
\]

Thus, there exists a unique nonzero monomial \( \bar{\lambda} \) of \( p \) such that

\[
\text{lm}(\bar{\lambda}(s_1, ..., s_n)) = \max \{ \text{lm}(\alpha(s_1, ..., s_n)) \mid \alpha \text{ is a monomial of } p \}.
\]

Since \( p(s_1, ..., s_n) = \sum \{ \alpha(s_1, ..., s_n) \mid \alpha \text{ is a monomial of } p \} \), we have \( \text{lm}(p(s_1, ..., s_n)) = \text{lm}(\bar{\lambda}(s_1, ..., s_n)) \neq 0 \), contradicting that \( p(s_1, ..., s_n) = 0. \)

3.7 Lemma. a) Let \((A, \leq)\) and \((B, \leq)\) be well ordered sets. Then \( A \times B \) is well ordered by the (lexicographic) order defined as \((a, b) \leq (a', b')\) if and only if \( a < a' \) or \((a = a' \text{ and } b \leq b')\).

b) In a well ordered set \((A, \leq)\) there exist no infinite strictly decreasing sequences.

c) For \( n \in \mathbb{N} \), the set \( T_n \) of the terms in \( R[X_1, ..., X_n] \) is well ordered by the lexicographic order (thus any strictly decreasing sequence of terms must be finite).

Proof. a) Recall that the ordered set \((A, \leq)\) is called well ordered if for any nonempty subset \( S \) of \( A \), \( \exists \alpha \in S \) such that \( \alpha \leq a, \forall a \in S \) (\( \alpha \) is unique with this property and is called the smallest element of \( S \). Thus, \( A \) is well ordered if any nonempty subset has a smallest element). Let \( \emptyset \neq S \subseteq A \times B \). Since \( S_1 := \{ a \in A \mid \exists b \in B \text{ cu } (a, b) \in S \} \neq \emptyset \), and \( A \) is well ordered, there exists its smallest element \( \alpha \in S_1 \) (so, \( \forall (a, b) \in S, \alpha \leq a \)). Let \( S_2 := \{ b \in B \mid (a, b) \in S \} \). There exists the smallest element \( \beta \) of \( S_2 \). Then \((\alpha, \beta)\) is the smallest element of \( S \): \( \forall (a, b) \in S \), we have \( \alpha < a \) (thus \((a, \beta) < (a, b)\) or \( \alpha = a \), in which case \( b \in S_2 \), so \( \beta \leq b \).

b) Let \((a_n)_{n \geq 1}\) be a decreasing sequence of elements in \( A \). Then the set \( \{ a_n \mid n \geq 1 \} \) has a smallest element \( a_k \). For any \( n \geq k \), we must have then \( a_k \leq a_n \); since \( a_n \leq a_k \) (the sequence is decreasing), \( a_n = a_k \) and the sequence is not strictly decreasing.
3. Symmetric polynomials

3.8 Corollary. (The fundamental theorem of the symmetric rational fractions) Let \( R \) be a domain and let \( K \) be its field of fractions. If \( p, q \in R[X_1, \ldots, X_n] \), \( q \neq 0 \), are such that \( p/q \) is a symmetric rational fractions, then there exist the polynomials \( f, g \in R[X_1, \ldots, X_n] \) such that
\[
\frac{p}{q} = \frac{f(s_1, \ldots, s_n)}{g(s_1, \ldots, s_n)}.
\]
In other words, the subfield of the rational symmetric fractions of the field \( K(X_1, \ldots, X_n) \) is \( K(s_1, \ldots, s_n) \).

**Proof.** If \( q \) is a symmetric polynomial, then \( p \) is symmetric, being the product \( q \cdot (p/q) \) in the subfield of the symmetric rational fractions. From 3.6, it follows that \( p, q \in R[s_1, \ldots, s_n] \). If \( q \) is not symmetric, let
\[
s = \prod_{\sigma \in S_n} \varphi_\sigma(q).
\]
Then \( s \) is symmetric and
\[
\frac{p}{q} = \frac{p \prod_{\sigma \in S_n} \varphi_\sigma(q)}{s},
\]
and we are in the conditions of the first case. \( \square \)

The symmetric polynomial \( t_m := X_1^m + \ldots + X_n^m \in R[X_1, \ldots, X_n] \) \((m \in \mathbb{N})\) is expressible using the fundamental symmetric polynomials \( s_1, \ldots, s_n \). The following identities allow a recursive computation of \( t_m \) as a polynomial of \( s_1, \ldots, s_n \).

3.9 Proposition. (Newton's identities) Let \( m \in \mathbb{N} \). In \( R[X_1, \ldots, X_n] \) the following relation holds:
\[
t_m = s_1 t_{m-1} - s_2 t_{m-2} + \ldots + (-1)^{m-2} s_{m-1} t_1 + (-1)^{m-1} ms_m.
\]
Proof. If \( m > n \), the convention \( s_k = 0 \) for \( k > n \) truncates the formula above (there are only \( n \) terms).

Let \( r \leq n \) and let \((a_1, \ldots, a_r)\) be an \( r \)-uple of natural numbers with \( a_1 \geq a_2 \geq \ldots \geq a_r \). Let \( s(a_1, \ldots, a_r) \) be the unique symmetric polynomial in \( R[X_1, \ldots, X_n] \) having the leading monomial \( X_1^{a_1} X_2^{a_2} \ldots X_r^{a_r} \).

For instance, \( s(m, 0, \ldots, 0) = X_1^m + \ldots + X_n^m = t_m \), \( s(1, 1, 0, \ldots, 0) = X_1X_2 + X_1X_3 + \ldots = s_2 \).

To simplify notations, let \( 1_i := (1, \ldots, 1) \) (1 appears \( i \) times) and \((a, 1_i) := (a, 1, \ldots, 1)\) (1 appears \( i \) times); also, we omit writing a sequence of 0's: \( s(m, 0, \ldots, 0) = s(m) \), \( s(1, 1, 0, \ldots, 0) = s(1, 1) = s_2 \), \( s(1_i, 0, \ldots, 0) = s(1_i) = s_i \). The following relations can be verified easily:

\[
\begin{align*}
    s_1 t_{m-1} &= t_m + s(m-1, 1) \\
    s_2 t_{m-2} &= s(m-1, 1) + s(m-2, 1, 1) \\
    s_3 t_{m-3} &= s(m-2, 1, 1) + s(m-3, 1, 1, 1) \\
    &\vdots
\end{align*}
\]

More generally, for any \( i \leq \min\{m-1, n\} \),

\[
s_i t_{m-i} = s(m-i+1, 1, i) + s(m-i, 1, i).
\]

If \( m \leq n \) and \( i = m-1 \), then

\[
s_{m-1} t_1 = s(2, 1, m-2) + ms_m.
\]

If \( m > n = i \), then

\[
s_n t_{m-n} = s(m-n+1, 1, n-1).
\]

Newton's identities follow by using the relations above in the sum

\[
\sum_{1 \leq i < m} (-1)^{i-1} s_i t_{m-i}. \quad \square
\]

4. Rings and modules of fractions

The method of construction of the field \( \mathbb{Q} \) from the domain \( \mathbb{Z} \) generalizes naturally to any commutative ring \( R \) (although in general
the result will not be a field). In $\mathbb{Q}$, all nonzero elements in $\mathbb{Z}$ become invertible. In many cases, there is no need that all nonzero elements in a ring $R$ become invertible in some “extension” of $R$. It is thus natural to define a concept corresponding to the notion of “set of denominators”. In what follows, all rings are commutative with identity and $R$ denotes such a ring.

4.1 Definition. A subset $S$ of $R$ is called a multiplicatively closed set (or a multiplicative set) if:

a) $1 \in S$.

b) $0 \notin S$.

c) $\forall s, t \in S \Rightarrow st \in S$.

These conditions are natural for a “set of denominators”: 1 must be a denominator, 0 cannot be one, and the product of two denominators is also a denominator.

For example, $\mathbb{Z} \setminus \{0\}$, $\mathbb{Z} \setminus 2\mathbb{Z}$, $\{2^n \mid n \in \mathbb{N}\}$ are multiplicatively closed sets in $\mathbb{Z}$.

For a given ring $R$ and a multiplicatively closed set $S \subseteq R$, we construct a ring $T$ and a ring homomorphism $\varphi : R \rightarrow T$, such that the images by $\varphi$ of all elements in $S$ are invertible in $T$ ($T$ is thus an $R$-algebra and $\varphi$ is its structural homomorphism).

4.2 Definition. On the set $R \times S = \{(a, s) \mid a \in R, s \in S\}$ we define the relation:

$$\forall (a, s), (b, t) \in R \times S, \text{ define: } (a, s) \sim (b, t) \iff \exists u \in S \text{ such that } u(ta - sb) = 0.$$ 

4.3 Remark. In the classic case of the construction of $\mathbb{Q}$ we have: $R = \mathbb{Z}$, $S = \mathbb{Z}^*$ and the following relation is used: $\forall (a, s), (b, t) \in R \times S$, $(a, s) \sim (b, t) \iff ta - sb = 0$. In this case the relation obtained coincides with the one in definition 4.2 (prove this!). The reason for adopting the
definition 4.2 is to handle the more general case when the multiplicatively closed set $S$ possibly contains zero divisors.

4.4 Proposition. The relation "~" is an equivalence relation on $R \times S$.

Proof. We prove only the transitivity. Let $(a, s), (b, t), (c, u) \in R \times S$ such that $(a, s) \sim (b, t)$ and $(b, t) \sim (c, u)$. Then $\exists v, w \in S$ such that $v(ta - sb) = 0$ and $w(ub - tc) = 0$. Multiply the first equality by $uw$ and the second by $sv$ and add. We obtain

$$uwvta - uwvsb + svwub - svwtc = 0 \Leftrightarrow vwt(ua - sc) = 0.$$  
Since $S$ is multiplicatively closed, $vwt \in S$, so $(a, s) \sim (c, u)$. □

4.5 Definition. Let $(a, s) \in R \times S$. The equivalence class of $(a, s)$ with respect to "~" is denoted by $\frac{a}{s}$ or $a/s$ and is called a fraction or a quotient (of denominator $s$ and numerator $a$). Thus:

$$\frac{a}{s} = a/s = \{(b, t) \in R \times S \mid (b, t) \sim (a, s)\}.$$  

The factor set $R \times S / ~$ (the set of all equivalence classes) is denoted by $S^{-1}R$:

$$S^{-1}R := \{a/s \mid a \in R, s \in S\}.$$  

It easy to see that $\frac{a}{s} = \frac{ta}{ts}$, $\forall s, t \in S$, $\forall a \in R$.

We define on $S^{-1}R$ two operations, having in mind the usual rules of addition and multiplication of fractions. For any $(a, s), (b, t) \in R \times S$, define:

$$\frac{a}{s} + \frac{b}{t} := \frac{ta + sb}{st}$$

$$\frac{a}{s} \cdot \frac{b}{t} := \frac{ab}{st}$$
4.6 Proposition. The operations defined above on $S^{-1}R$ are correctly defined and $S^{-1}R$ becomes a commutative ring with identity. The elements $0$ and $1$ in $S^{-1}R$ are:

\[
0 = \frac{0}{1} = \frac{s}{s}, \forall s \in S;
\]

\[
1 = \frac{1}{1} = \frac{s}{s}, \forall s \in S.
\]

The mapping $\varphi : R \to S^{-1}R$, $\varphi(a) = a/1$, $\forall a \in R$, is a ring homomorphism, called the canonical homomorphism (thus $S^{-1}R$ is an $R$-algebra).

\textbf{Proof.} We check that the addition is correctly defined. Let $(a, s), (b, t), (a', s'), (b', t') \in R \times S$, such that $(a, s) \sim (a', s')$ and $(b, t) \sim (b', t')$. We must show that $(ta + sb, st) \sim (t'a' + s'b', s't')$. Let $u, v \in S$ such that $u(s'a - sa') = 0$ and $v(t'b - tb') = 0$. Multiply the first of these equalities by $tt'v$ and the second by $ss'u$ and add them. We obtain

\[
vu((ta + sb)s't' - (t'a' + s'b')st) = 0.
\]

The rest of the proof (the multiplication is correctly defined; checking the axioms for the ring $S^{-1}R$) is left to the reader. 

Note that any $s \in S$ is taken by $\varphi$ into an invertible element in $S^{-1}R$: $\varphi(s) = s/1$ has the inverse $1/s \in S^{-1}R$.

Moreover, the homomorphism $\varphi$ is injective $\iff$ $S$ contains no zero divisors. Indeed, $a/1 = 0/1 \iff \exists u \in S$ such that $ua = 0$.

If $0 \in S$, then $S^{-1}R$ is the zero ring (with only one element, $0/1 = a/s$, $\forall a \in R$, $\forall s \in S$); for this reason the condition $0 \not\in S$ is imposed in the definition of a multiplicatively closed set.

The ring $S^{-1}R$ is called the ring of fractions (or the ring of quotients) of $R$ with respect to the multiplicatively closed set $S$.

In the important case when $R$ is a domain and $S = R \setminus \{0\}$, $S^{-1}R$ is a field, called the field of fractions (or field of quotients) of $R$, and denoted $Q(K)$. 

4.7 Example. a) The field of quotients of $\mathbb{Z}$ is $\mathbb{Q}$.

b) For any field $K$, the field of quotients of the polynomial ring $K[X]$ is denoted by $K(X)$ and is called the field of rational fractions with coefficients in $K$. Its elements are “fractions” of the form $\frac{f}{g}$, where $f, g \in K[X]$, with $g \neq 0$. Similarly, the field of quotients of the polynomial ring $K[X_1, \ldots, X_n]$ is denoted by $K(X_1, \ldots, X_n)$ and is called the field of rational fractions in $n$ indeterminates with coefficients in $K$.

c) The field of quotients of $\mathbb{Z}[X]$ is (isomorphic to) $\mathbb{Q}(X)$ (prove this!).

For any domain $R$, $Q(R)$ is the “smallest” field that “includes” $R$. More generally, $S^{-1}R$ is the “smallest” ring that includes $R$ (if $S$ contains no zero divisors) such that all elements in $S$ are invertible in $S^{-1}R$. This fact is stated rigorously as follows:

4.8 Theorem. (The universality property of the ring of fractions)
Let $S$ be a multiplicatively closed subset of $R$. Then $S^{-1}R$ is a commutative ring with identity and $\varphi : R \to S^{-1}R$ is a ring homomorphism such that $\varphi(s)$ is invertible in $S^{-1}R$, $\forall s \in S$. Moreover, the pair $(\varphi, S^{-1}R)$ is universal relative to this property, namely:

For any pair $(\gamma, T)$ where $T$ is a commutative ring with identity and $\gamma : R \to T$ is a ring homomorphism, such that $\gamma(s)$ is invertible in $T$, $\forall s \in S$, there exists a unique ring homomorphism $g : S^{-1}R \to T$ such that $\gamma = g\varphi$.

Proof. Define $g(a/s) = \gamma(a)(\gamma(s))^{-1}$, $\forall a \in R$, $\forall s \in S$. The reader can easily verify that $g$ is correctly defined, that it is a ring homomorphism and it is the only one with $\gamma = g\varphi$.  \[\square\]
The “complete” structure of $S^{-1}R$ is that of $R$-algebra, the canonical homomorphism $\phi$ being the structural homomorphism. In this setting, the above property reads:

For any commutative $R$-algebra $(\gamma, T)$, where $\gamma: R \to T$ is the structural homomorphism, such that $\gamma(s)$ is invertible in $T$ for any $s \in S$, there exists a unique $R$-algebra homomorphism $g : S^{-1}R \to T$.

As expected, the universality property of the ring of fractions determines the ring of fractions up to a (unique) isomorphism:

4.9 Theorem. Let $S$ be a multiplicatively closed subset of $R$. Assume $B$ is a commutative ring with identity and $\beta : R \to B$ is a homomorphism satisfying the property:

For any commutative ring with identity $T$ and any ring homomorphism $\gamma : R \to T$ such that $\gamma(s)$ is invertible in $T$, $\forall s \in S$, there exists a unique ring homomorphism $g : B \to T$ such that $\gamma = g\beta$.

Then there exists a unique ring isomorphism $h : S^{-1}R \to B$ such that $h\phi = \beta$. □

The above construction can be applied to an $R$-module $M$, with minor modifications. Given a multiplicatively closed subset $S \subseteq R$ and an $R$-module $M$, we define on $M \times S$ the equivalence relation: $\forall (a, s), (b, t) \in M \times S, (a, s) \sim (b, t) \iff \exists u \in S$ such that $u(ta - sb) = 0$ (cf. 4.2). The following result is proven exactly like in the case of $S^{-1}R$:

4.10 Proposition. Let $M$ be an $R$-module and let $S$ be a multiplicatively closed subset of $R$. Then the relation "~" defined above is an equivalence relation on $M \times S$. Denoting by $\frac{x}{s}$ the equivalence class of $(x, s) \in M \times S$ and $S^{-1}M := \left\{ \frac{x}{s} \mid x \in M, s \in S \right\}$, $S^{-1}M$ becomes an Abelian group with the addition: $\forall x, y \in M, \forall s, t \in S,$
Moreover, \( S^{-1}M \) is an \( S^{-1}R \)-module with the multiplication defined by: \( \forall a \in R, \forall x \in M, \forall s, t \in S, \)

\[
\frac{a}{s} \cdot \frac{x}{t} := \frac{ax}{st}.
\]

The \( S^{-1}R \)-module \( S^{-1}M \) is called the module of fractions (or quotients) of \( M \) relative to the multiplicatively closed subset \( S \). The homomorphism \( \phi_M : M \to S^{-1}M, \phi_M(x) = x/1, \forall x \in M, \) is called the canonical homomorphism.

The connection between the ideals of \( R \) and the ideals of the ring of fractions is very close. An immediate property is:

**4.11 Proposition.** Let \( I \) be an ideal in the ring \( R \). Then \( S^{-1}I := \{a/s \mid a \in I, s \in S\} \) is an ideal in \( S^{-1}R \). Moreover, any ideal in \( S^{-1}R \) is of the form \( S^{-1}I \), for some ideal \( I \) in \( R \).

**Proof.** It is immediate that \( S^{-1}I \leq S^{-1}R \) if \( I \leq R \). If now \( J \leq S^{-1}R \), let \( I := \phi^{-1}(J) \) (an ideal in \( R \)). We have \( a/1 \in J \iff \exists \ s \in S \) such that \( a/s \in J \). Thus \( \phi^{-1}(J) = \{a \in R \mid \exists s \in S \text{ such that } a/s \in J\} \). Then \( S^{-1}I = \{a/s \mid a \in I, s \in S\} = J \).

A similar connection exists between the submodules of \( M \) and the submodules of \( S^{-1}M \) (can you state and prove it?).

**4.12 Definition.** A multiplicatively closed subset \( S \subseteq R \) is called saturated if all the divisors of the elements in \( S \) are also in \( S \): \( \forall s \in S, \forall d, r \in R, dr = s \) implies \( d \in S \) and \( r \in S \). If \( S \) is an arbitrary multiplicatively closed set, let

\[
S' := \{d \in R \mid \exists r \in R, \exists s \in S \text{ such that } dr = s\}.
\]

\( S' \) is called the saturate of the multiplicative set \( S \). Evidently, \( S \) is saturated \( \iff \) \( S = S' \).
4. Rings and modules of fractions

The following property says that the any ring of fractions can be constructed using a *saturated* multiplicative set.

**4.13 Proposition.** Let $S$ be multiplicative set in the ring $R$. Then:

a) $S'$ is a saturated multiplicative set of $R$.

b) There exists a canonical isomorphism $S^{-1}R \cong S'^{-1}R$.

**Proof.**
a) Check the definition.

b) We denote the equivalence relation on $S' \times R$ (defined as in 4.2) with $\approx$; the equivalence class of $(a, s)$ in $R \times S'$ is denoted by $a//s \in S'^{-1}R$ (in order to distinguish from the fraction $a/s$ which may be in $S^{-1}R$). Define the canonical mapping $\varphi: S^{-1}R \to S'^{-1}R$, $\varphi(a/s) = a//s, \ \forall a/s \in S^{-1}R$. The definition is independent on the choice of the representatives $a$ and $s$: if $(a, s) \sim (b, t)$, then $(a, s) \approx (b, t)$. Clearly, $\varphi$ is a ring homomorphism. We have $\text{Ker} \varphi = \{a/s \in S^{-1}R \mid a//s = 0/1\}$. But $a//s = 0//1 \iff \exists u \in S'$ such that $ua = 0$. Thus, $\exists r \in R$ such that $ur \in S$ and $ura = 0$, i.e. $a/s = 0/1$. Therefore, $\text{Ker} \varphi = \{0/1\}$. Also, $\varphi$ is surjective: if $a//d \in S'^{-1}R$, with $a \in R$, $d \in S'$, then there exists $r \in R$ such that $dr \in S$. It is clear then that $r \in S'$, so $a//d = ar//dr = \varphi(ar/dr)$.

An important example of a multiplicative set and its corresponding ring of fractions is the following:

**4.14 Proposition.** Let $P$ be a prime ideal in the ring $R$. Then $S := R \setminus P$ is a multiplicative subset in $R$ and the ring of fractions $S^{-1}R$ has a unique maximal ideal (it is a local ring).

**Proof.** The definition of a prime ideal can be stated as follows: for any $a$, $b \not\in P$, then $ab \not\in P$, which means that $S$ is multiplicatively closed. If $I \leq R$, with $I \cap S \neq \emptyset$, then $S^{-1}I = S^{-1}R$. Indeed, if $s \in I \cap S$, then $s/1 \in S^{-1}I$ and it is invertible, so $S^{-1}I = R$. Thus, the proper ideals of $S^{-1}R$ are of the form $J = S^{-1}I$, where $I \cap S = \emptyset$ ($\iff I \subseteq P$), so $J \subseteq S^{-1}P$. But $S^{-1}P$ is a proper ideal: if $1/1 = p/s$, with $p \in P$, $s \in S$, then
then \( \exists u \in S \) such that \( u(s - p) = 0 \Rightarrow us \in P \Rightarrow u \in P \) or \( s \in P \), contradicting \( S = R \setminus P \). Thus, \( S^{-1}P \) is the unique maximal ideal in \( S^{-1}R \). \( \square \)

If \( u : M \to N \) is an \( R \)-module homomorphism, then we define the mapping

\[
S^{-1}u : S^{-1}M \to S^{-1}N, (S^{-1}u)(x/s) := u(x)/s, \forall x \in M, \forall s \in S.
\]

\( S^{-1}u \) is easily seen to be an \( S^{-1}R \)-module homomorphism and it is the unique \( S^{-1}R \)-module homomorphism with the property that \( (S^{-1}u) \circ \varphi_M = \varphi_N \circ u \), where \( \varphi_M : M \to S^{-1}M \) and \( \varphi_N : N \to S^{-1}N \) are the canonical homomorphisms.

Thus, for a fixed multiplicatively closed subset \( S \), we defined a functor:

\[
S^{-1} - : R-Mod \to S^{-1}R-Mod.
\]

Moreover, \( S^{-1} - \) is an additive functor: for any \( R \)-module homomorphisms \( u_1, u_2 : M \to N \),

\[
S^{-1}(u_1 + u_2) = S^{-1}u_1 + S^{-1}u_2.
\]

4.15 Proposition. The functor \( S^{-1} - : R-Mod \to S^{-1}R-Mod \) is exact: if the sequence

\[
A \xrightarrow{u} B \xrightarrow{v} C
\]

is exact in \( R-Mod \), then the sequence

\[
S^{-1}A \xrightarrow{S^{-1}u} B \xrightarrow{S^{-1}v} S^{-1}C
\]

is exact in \( S^{-1}R-Mod \). \( \square \)

The unproven statements above are proposed as exercises.

5. Categories, functors

The category language is nowadays all-pervading throughout mathematics. Introduced in 1942 by MacLane and Eilenberg, the con-
cept of category marks a new step of abstraction in mathematics. For instance, from the abstract notion of integer one passes to the notion of set of integers $\mathbb{Z}$ (as a new object of study, endowed with a certain structure). Generalizing key properties of $\mathbb{Z}$ yields the concept of ring. This leads to the study of structures given by axioms (such as the structures of group, ring, field, topological space etc). For a certain type of structure, usually a natural notion of homomorphism arises (for example, the familiar notion of group homomorphism, or ring homomorphism).

The philosophy in the category theory is to study the class of all structures of a certain type (for instance the class of all rings) using the homomorphisms between these structures, and ignoring the elements of these structures. One advantage of this approach is given by generality: a result that holds in any category is valid in the category of groups, and also in the category of topological spaces etc. Besides many results and clarifications brought in almost all areas of mathematics, the category theory simplifies, unifies (to a certain extent) and standardizes the language of mathematics.

We present here some basic concepts on categories that are useful for a better understanding of several topics. A detailed presentation of category theory is to be found for instance in Herrlich, Strecker [1979].

Before proceeding to the definition of a category, we briefly describe the concept of class.

The notion of class is introduced in the axiomatic set theory (in order to avoid the paradoxes generated by considering “very large sets”). In this theory, the notion of set and the relation "$\in$" ("belongs to") are primary notions (are undefined); also, all objects are sets. Any element of a set is thus also a set.
Formally, (in the Zermelo-Fraenkel theory\textsuperscript{84}) a \textit{class} is an expression of the formal language of the set theory\textsuperscript{85} that contains exactly one free variable (in other words, a predicate with one variable). Of course, if \(P(x)\) is a predicate, the class \(P\) defined by \(P(x)\) is intuitively the “collection” of all objects (i.e. sets) \(x\) for which \(P(x)\) is true. Instead of writing "\(P(x)\) is true" one writes "\(x \in P\)" by analogy with the set language. Any set \(A\) is a class (corresponding to the predicate "\(x \in A\)"), but there exist classes that are not sets (for instance, the class of all sets, defined by the predicate "\(x = x\)").

If \(P\) and \(Q\) are classes, defined by the predicates \(P(x)\) and \(Q(x)\), their \textit{union} \(P \cup Q\) is the class \(P(x) \lor Q(x)\); their \textit{intersection} \(P \cap Q\) is the class \(P(x) \land Q(x)\). We say that \(P \subseteq Q\) if the proposition \(\forall x(P(x) \rightarrow Q(x))\) is true. Similarly, one can define the analogue for classes of the usual set operations. For details, see

5.1 \textbf{Definition}. A \textit{category} \(\mathcal{C}\) consists of the following data:

- a \textit{class} \(\text{Ob } \mathcal{C}\). The elements of \(\text{Ob } \mathcal{C}\) are called the \textit{objects} of the category \(\mathcal{C}\).

- for every couple \((A, B)\), where \(A, B \in \text{Ob } \mathcal{C}\), a (possibly empty) set \(\text{Hom}_\mathcal{C}(A, B)\) is given. The elements of \(\text{Hom}_\mathcal{C}(A, B)\) are called \textit{morphisms} (or \textit{arrows}) from \(A\) to \(B\). The fact that \(u \in \text{Hom}_\mathcal{C}(A, B)\) is also written \(u : A \rightarrow B\) or \(\xymatrix{A \ar[r]^-u & B}\); \(A\) is the \textit{domain} (or \textit{source}) and \(B\) is

\textsuperscript{84} In the Gödel-Bernays-von Neumann theory, the notion of class is a primary notion. The sets are exactly the classes that are elements of some class.

\textsuperscript{85} We do not define here this formal language. Roughly speaking, it consists of expressions (strings of symbols) formed from the \textit{atomic} expressions (of the type \(x = y\) or \(x \in y\)) by using the \textit{logical operators} \(\lor, \land, \neg\) and the \textit{quantifiers} \(\forall\) and \(\exists\). For instance "\((\forall x(x \in y)) \land (\exists z(y = z) \lor \neg(z = a))" is an expression, in which \(x\) and \(z\) are \textit{bound} variables and \(y\) is a \textit{free} variable. If \(a\) is assumed to be a constant, then this expression is a \textit{predicate} (it has one free variable, namely \(y\)). An expression with no free variables is a \textit{proposition}. 
the codomain (or sink) of $u$. The class $\bigcup \{\text{Hom}_C(A, B) \mid A, B \in \text{Ob } C\}$ is denoted $\text{Hom } C$ and is called the class of the morphisms of the category $C$.\footnote{Formally, it is the class defined by the predicate $H(u) = \exists A \exists B (A \in \text{Ob } C \land B \in \text{Ob } C \land u \in \text{Hom}(A, B))$.}

- for any triple $(A, B, C)$ of objects of $C$ there exists a function defined on $\text{Hom}_C(B, C) \times \text{Hom}_C(A, B)$ with values in $\text{Hom}_C(A, C)$. The image of the couple $(v, u)$ is denoted by $v \circ u$ (or simply $vu$) and is called the composition of the morphisms $v$ and $u$.

In any category $C$ the following axioms must be satisfied:

1) Any morphism has a unique domain and a unique codomain: for any $A, B, C, D \in \text{Ob } C$, $(A, B) \neq (C, D)$ implies

$$\text{Hom}_C(A, B) \cap \text{Hom}_C(C, D) = \emptyset.$$

2) The composition of morphisms is associative: $\forall A, B, C, D \in \text{Ob } C$ and $\forall u : A \to B$, $v : B \to C$, $w : C \to D$, we have $w(uv) = (wu)v$ (denoted usually by $wuv$).

3) $\forall A \in \text{Ob } C$, there exists a morphism $1_A : A \to A$ (called the identity morphism of $A$) such that, $\forall B \in \text{Ob } C$ and $\forall u : A \to B$, $v : B \to A$, we have $u \circ 1_A = u$ and $1_A \circ v = v$.

5.2 Remark. a) The identity morphism of an object $A$ is unique: if $j : A \to A$ is an identity morphism of $A$, then $j = j \circ 1_A = 1_A$.

b) In the definition of a category, the morphisms are essential. One can identify an object $A \in \text{Ob } C$ with its identity morphism $1_A$. The notion of category can be defined using only the concept of morphism.

5.3 Examples. a) The category $\text{Set}$ of all sets. Its objects are sets. If $A$ and $B$ are sets, $\text{Hom}_{\text{Set}}(A, B)$ is the set of all functions $\varphi : A \to B$. The composition of morphisms in $\text{Set}$ is the usual function composition. The identity morphism of $A$ is the identity function of $A$. 
b) The category $Gr$ of groups. $\text{Ob} \ Gr$ is the class of groups and $\text{Hom}_{Gr}(G, H)$ is the set of group homomorphisms from $G$ to $H$, $\forall G, H \in \text{Ob} \ Gr$. As in $Set$, the composition is the usual function composition.

c) Let $R$ be a ring with identity. The category $R-\text{Mod}$ has as objects left $R$-modules, and the morphisms are $R$-module homomorphisms, with the usual composition. In the same way one defines the category $\text{Mod}-R$ of right $R$-modules.

d) One can define similarly the following categories:
- $\text{Ring}$: the rings with the ring homomorphisms.
- $\text{Ring}_u$: the rings with identity, the morphisms are ring homomorphism preserving the identity.
- $\text{Ab}$: the Abelian groups, with the group homomorphisms.
- $\text{Poset}$: the (partially) ordered sets, the morphisms being the order preserving mappings.

d) Let $(A, \leq)$ be a set equipped with a preorder relation (a transitive and reflexive relation). Define a category $\mathcal{A}$, with $\text{Ob} \ A := A$. For any $a, b \in \text{Ob} \ A$, set
$$\text{Hom}_A(a, b) = \begin{cases} \{ (a, b) \}, & \text{if } a \leq b \\ \emptyset, & \text{else} \end{cases}$$

The reader is invited to define the composition of morphisms and check the axioms 1) – 3).

e) Let $(G, \cdot)$ be a monoid (the operation is associative and has a neutral element). Define a category $\mathcal{G}$ as follows: $\text{Ob} \ G$ is a set with one element (for instance $\text{Ob} \ G = \{ G \}$), and $\text{Hom}_G(G, G) = G$ (the morphisms are the elements of $G$); the composition of the morphisms $a, b \in G$ is $a \cdot b$ (where $\cdot$ is the operation on $G$). The identity morphism is the identity element.

The reader can easily produce other examples of categories, based on hers/his mathematical background (the category of semigroups, the category of finite sets, the category of fields, the category of topologi-
cal spaces, the continuous mapping being the morphisms etc.). In each situation it is necessary to state exactly the class of the objects of the category, the set of morphisms between two arbitrary objects, the composition of morphisms and check axioms 1)-3).

Often the writing $A \in \mathcal{C}$ replaces $A \in \text{Ob} \mathcal{C}$, if no confusion arises.

5.4 Definition. A category $\mathcal{C}$ is called a subcategory of a category $\mathcal{D}$ if $\text{Ob} \mathcal{C} \subseteq \text{Ob} \mathcal{D}$ and, $\forall A, B \in \text{Ob} \mathcal{C}$, $\text{Hom}_\mathcal{C}(A, B) \subseteq \text{Hom}_\mathcal{D}(A, B)$; moreover, the composition of two morphisms in $\mathcal{C}$ is their composition in $\mathcal{D}$. We call $\mathcal{C}$ a full subcategory of $\mathcal{D}$ if $\forall A, B \in \text{Ob} \mathcal{C}$, $\text{Hom}_\mathcal{C}(A, B) = \text{Hom}_\mathcal{D}(A, B)$.

For instance, $\text{Ab}$ is a full subcategory of $\text{Gr}$; $\text{Gr}$ is a subcategory of $\text{Set}$, but not a full subcategory.

5.5 Definition. We define now some remarkable objects and morphisms in a category $\mathcal{C}$ that generalize familiar concepts.

a) An object $I \in \mathcal{C}$ is called an initial object if $\forall A \in \mathcal{C}$, $|\text{Hom}_\mathcal{C}(I, A)| = 1$ (there exists a unique morphism $I \rightarrow A$). An object $F$ is called a final object if $\forall A \in \mathcal{C}$, $|\text{Hom}_\mathcal{C}(A, I)| = 1$. An object that is simultaneously initial and final is called a zero object.

b) A morphism $u : A \rightarrow B$ is called:

- a monomorphism if $\forall C \in \mathcal{C}$, $\forall v, w \in \text{Hom}_\mathcal{C}(B, A)$, $uv = uw$ implies $v = w$.

- an epimorphism if $\forall C \in \mathcal{C}$, $\forall v, w \in \text{Hom}_\mathcal{C}(B, C)$, $vu = wv$ implies $v = w$.

- a bimorphism if it is both a monomorphism and an epimorphism.

- an isomorphism if there exists $v : B \rightarrow A$ such that $uv = 1_B$ and $vu = 1_A$ (v is called then the inverse of u). The notation $A \xrightarrow{\sim} B$ is used to denote an isomorphism.
Two objects \( A, B \in \mathcal{C} \) are called isomorphic (written \( A \cong B \)) if there exists an isomorphism \( A \to B \). The relation of isomorphism on the class \( \text{Ob} \mathcal{C} \) is an equivalence relation.

5.6 Examples. a) In \( Gr \) there exist initial objects, namely the groups having one element (necessarily the neutral element of that group). These are also final objects (thus they are zero objects) in \( Gr \). The same remark holds for \( Ab \) and \( R\text{-Mod} \).

b) In \( Set \) the empty set \( \emptyset \) is the only initial object\(^{87}\); any set with one element is a final object. \( Set \) has no zero objects.

c) The monomorphisms in \( Set \) (as in \( Gr, Ab, R\text{-Mod} \)) are the morphisms that are injective functions. Which are the epimorphisms? In these categories the isomorphisms coincide with the bimorphisms.

d) In the category \( Ann \) of rings with identity, the inclusion \( \mathbb{Z} \to \mathbb{Q} \) is a monomorphism and an epimorphism, and is not a surjective function or an isomorphism.

5.7 Proposition. Let \( \mathcal{C} \) be a category and let \( A, B \) be initial objects in \( \mathcal{C} \). Then there exists a unique isomorphism \( A \cong B \).

Proof. Since \( A \) is an initial object, there exists a unique morphism \( \varphi : A \to B \). But \( B \) is an initial object, thus there exists a unique morphism \( \psi : B \to A \). The morphism \( \psi\varphi : A \to A \) is equal to \( 1_A \) (because there exists a unique morphism \( A \to A \)). Likewise, \( \varphi\psi = 1_B \). So, \( \varphi \) and \( \psi \) are isomorphisms, inverse to each other. \( \square \)

It is important to point out that various “universality properties” that some objects satisfy are simply a restatement of the fact that those objects are initial (or final) objects in certain categories. In this situation, the propositions of the type “the universality property of ... determines ... up to a unique isomorphism” merely translate for a spe-

\(^{87}\) For any set \( A \), there exists a unique function \( \emptyset \to A \), namely the function \( \emptyset \).
cific category the assertion “between any two initial (final) objects in a category there exists a unique isomorphism”.

5.8 Examples. a) The direct sum. Let \((M_i)_{i \in I}\) be a family of objects in \(R\text{-Mod}\). Let \(S\) be the category whose objects are couples of the form \((S, (\sigma_i)_{i \in I})\), where \(S \in R\text{-Mod}\) and \(\sigma_i : M_i \to S\) are morphisms in \(R\text{-Mod}\), \(\forall i \in I\). If \((S, (\sigma_i)_{i \in I}), (T, (\tau_i)_{i \in I}) \in S\), define the morphisms \(\varphi : S \to T\) such that \(\varphi \sigma_i = \tau_i\), \(\forall i \in I\). Check the axioms for a category and the following assertion:

\((S, (\sigma_i)_{i \in I})\) is a direct sum of the family \((M_i)_{i \in I}\), of canonical injections \((\sigma_i)_{i \in I}\), is tantamount to saying that \((S, (\sigma_i)_{i \in I})\) is an initial object in \(S\).

b) The direct product. Let \((M_i)_{i \in I}\) be a family of objects in \(R\text{-Mod}\). Then: \((P, (\pi_i)_{i \in I})\) is a direct product of the family \((M_i)_{i \in I} \leftrightarrow (P, (\pi_i)_{i \in I})\) is a final object in a certain category (describe it!).

An important principle, often invoked, is the principle of duality.

5.9 Definition. If \(\mathcal{C}\) is a category, the dual category \(\mathcal{C}^\circ\) is defined as follows: \(\text{Ob} \mathcal{C}^\circ := \text{Ob} \mathcal{C}\); if \(A \in \text{Ob} \mathcal{C}\), let \(A^\circ\) be the object \(A\) seen in \(\mathcal{C}^\circ\). For any \(A, B \in \text{Ob} \mathcal{C}\), let \(\text{Hom}_{\mathcal{C}^\circ}(B^\circ, A^\circ) := \text{Hom}_{\mathcal{C}}(A, B)\). A morphism \(u : A \to B\) in \(\mathcal{C}\) is denoted \(u^\circ : B^\circ \to A^\circ\) in \(\mathcal{C}^\circ\). The Composition in \(\mathcal{C}^\circ\) of the morphisms \(u^\circ : B^\circ \to A^\circ\) and \(v^\circ : C^\circ \to B^\circ\) is defined by \(u^\circ v^\circ := (vu)^\circ\), where \(vu\) is the composition of \(u : A \to B\) and \(v : B \to C\) in \(\mathcal{C}\). Evidently, there exists \(1_{A^\circ} = (1_A)^\circ\).

Intuitively, the dual of the category \(\mathcal{C}\) is obtained by “reversing the arrows” in \(\mathcal{C}\) (and reversing the order of composing the arrows).

Let \(P\) be a statement formulated in terms of objects and morphisms. For each category \(\mathcal{C}\), one obtains a proposition, denoted \(P(\mathcal{C})\). Let \(P^\circ\)
be the dual statement (obtained from $P$ by reversing the arrows and the order in composing the arrows)\textsuperscript{88}. The principle of duality is the following: If $P$ is valid in any category, then $P^\circ$ is valid in any category.

Similarly, any notion (definition) in a category has a dual notion, obtained by reversing the arrows and the order in composing the arrows. A notion that coincides with its dual is called autodual.

**5.10 Example.** a) The dual of the notion of initial object is the notion of final object.

b) The dual of the notion of monomorphism is the notion of epimorphism.

c) The notion of isomorphism is autodual.

d) We saw that: for any category $\mathcal{C}$, any two initial objects in $\mathcal{C}$ (if any) are isomorphic. By dualization, one obtains (no new proof needed): for any category $\mathcal{C}$, any two final objects in $\mathcal{C}$ (if any) are isomorphic.

The intuitive concept of “morphism of categories” is the notion of functor.

**5.11 Definition.** Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A covariant functor $F$ from $\mathcal{C}$ to $\mathcal{D}$, denoted $F : \mathcal{C} \to \mathcal{D}$, is a couple $F = (F', F'')$, where $F' : \text{Ob } \mathcal{C} \to \text{Ob } \mathcal{D}$, $F'' : \text{Hom } \mathcal{C} \to \text{Hom } \mathcal{D}$, such that:

(F1) $\forall A, B \in \text{Ob } \mathcal{C}$, $F''(\text{Hom}_\mathcal{C}(A, B)) \subseteq \text{Hom}_\mathcal{D}(F'(A), F'(B))$; in other words, if $u : A \to B$, then $F''(u) : F'(A) \to F'(B)$.

(F2) $F$ preserves the composition of morphisms: $\forall A, B, C \in \text{Ob } \mathcal{C}$ and $\forall u : A \to B$, $v : B \to C$, then $F''(v \circ u) = F''(v) \circ F''(u)$.

(F3) $F$ preserves the identity morphisms: $\forall A \in \mathcal{C}$, $F''(1_A) = 1_{F'(A)}$.

---

\textsuperscript{88} in other words, $P^\circ(\mathcal{C})$ is the same thing as $P(\mathcal{C}^\circ)$ interpreted in $\mathcal{C}$.
A contravariant functor $F : \mathcal{C} \to \mathcal{D}$ satisfies F3 and the duals of F1 and F2:

\begin{align*}
(F1^*) \quad & F''(\text{Hom}_C(A, B)) \subseteq \text{Hom}_D(F''(B), F'(A)), \forall A, B \in \text{Ob} \mathcal{C}.
(F2^*) \quad & \forall A, B, C \in \text{Ob} \mathcal{C} \text{ and } \forall u : A \to B, \quad v : B \to C, \text{ then } F''(v \circ u) = F''(u) \circ F''(v).
\end{align*}

A contravariant functor “reverses the arrows”. A contravariant functor from $\mathcal{C}$ to $\mathcal{D}$ is the same as a covariant functor from $\mathcal{C}$ to $\mathcal{D}^\circ$ (or from $\mathcal{C}^\circ$ to $\mathcal{D}$). This is the reason why the results on covariant functors transfer to contravariant functors. In what follows, “functor” means “covariant functor”.

The distinction between the components $F'$ and $F''$ of the functor $F$ is usually dropped, denoting $F(A)$ instead of $F'(A)$ and $F(u)$ instead of $F''(u)$. Besides, $F$ is perfectly determined by $F''$, its action on morphisms, cf. 5.2).

A functor $F : \mathcal{C} \to \mathcal{D}$ is called:

- faithful if $\forall A, B \in \mathcal{C}$, $F_{A,B} : \text{Hom}_C(A, B) \to \text{Hom}_D(FA, FB)$ is an injective function.
- full if $\forall A, B \in \mathcal{C}$, $F_{A,B} : \text{Hom}_C(A, B) \to \text{Hom}_D(FA, FB)$ is a surjective function.
- fully faithful if it is full and faithful.

We are interested in categories in which the morphisms are functions between sets:

**5.12 Definition.** The category $\mathcal{C}$ is called **concrete** if there exists a covariant faithful functor $F : \mathcal{C} \to \text{Set}$.

The category $\text{Gr}$ is concrete: define the functor $U : \text{Gr} \to \text{Set}$, by associating to any group its underlying set\(^{89}\) and sending any group

---

\(^{89}\) A group $G$ is, formally, a **couple** $(G, \cdot)$, where $G$ is the underlying set of the group and $\cdot : G \times G \to G$ is the group operation. Thus, a function from a group $(G, \cdot)$
homomorphism $u$ to $u$, seen as a function between the underlying sets. Then $U$ is a faithful functor (which is not full). $U$ is called a “forgetful functor”: it “forgets” the group structure. Similarly, $Ab$, $Ring$, $R$-mod are concrete categories (why?).

In concrete categories there is a notion of free object on a set $X$ (compare the following definition with the universality property of the free $R$-module on a set $X$).

**5.13 Definition.** Let $\mathcal{C}$ be a concrete category and let $F: \mathcal{C} \to Set$ be a covariant faithful functor. The object $L$ in $\mathcal{C}$ is called free (relative to $F$) over the set $X \subseteq F(L)$ if, for any $A \in \mathcal{C}$ and any function $\gamma: X \to F(A)$, there exists a unique morphism in $\mathcal{C}$, $g: L \to A$, such that $F(g)|_X = \gamma$.

If $\mathcal{C}$ is one of the categories $Gr$, $Ab$, $R$-mod, $Ring$, and $F$ is the forgetful functor, the definition reads: $L \in \mathcal{C}$ is free over the set $X \subseteq L$ if, for any object $A \in \mathcal{C}$ and any function $\gamma: X \to A$, there exists a unique morphism in $\mathcal{C}$, $g: L \to A$, such that $g|_X = \gamma$.

The following notion allows comparing two functors and is the analogue of the concept of homomorphism of algebraic structures.

**5.14 Definition.** Let $F, G: \mathcal{C} \to \mathcal{D}$ be covariant functors. A natural transformation (or a functor morphism) $\alpha: F \to G$ is given if for any object $A \in \mathcal{C}$, there exists a morphism in $\mathcal{D}$, $\alpha_A: F(A) \to G(A)$, such that, for any $u: A \to B$ morphism in $\mathcal{C}$, $\alpha_B \circ F(u) = G(u) \circ \alpha_A$, i.e. the following diagram (of morphisms in $\mathcal{D}$) is commutative:

---

to a group $(H, \ast)$ is not the same thing as a function between their underlying sets $G$ and $H$. For this reason, $Gr$ is not a subcategory of $Set$. 

6. Solvable groups

The notion of solvable group is closely connected with the solvability by radicals of a polynomial.

In what follows, \((G, \cdot)\) is a group and its neutral element is denoted by 1. The trivial subgroup \(\{1\}\) is also denoted 1. The notation \(H \leq G\) signifies “\(H\) is a subgroup of \(G\)”, and \(H \triangleleft G\) means “\(H\) is a normal subgroup of \(G\)”.

6.1 Definition. The group \(G\) is called a solvable group if there exists a finite chain of subgroups of \(G\):

\[
1 = G_0 \leq G_1 \leq \ldots \leq G_n = G, \quad (S)
\]

such that:

i) \(G_{i-1} < G_i, \forall i, 1 \leq i \leq n;\)

ii) The factor groups \(G_i/G_{i-1}\) are Abelian, \(\forall i, 1 \leq i \leq n.\)
A chain of subgroups \((S)\) that satisfies (i) is called a *normal series of* \(G\)\(^{90}\). A chain \((S)\) satisfying i) and ii) is called a *solvable series of* \(G\). The number \(n\) is called the *length* of the series \((S)\). The groups \(G_i/G_{i-1}\) are called the *factors* of the normal series \((S)\).

**6.2 Examples.** a) Any Abelian group is solvable.

b) Let \(n \in \mathbb{N}^*\) and let \(S_n\) be the group of all permutations on \(n\) objects (also called the *symmetric group on* \(n\) *objects*). Let \(A_n\) be the *alternating group on* \(n\) *objects*, namely the subgroup of \(S_n\) formed by the even permutations. \(A_n\) is a subgroup of index 2 in \(S_n\) (it is the kernel of the *sign homomorphism* \(\varepsilon: S_n \rightarrow \{-1, 1\}\)), thus it is normal in \(S_n\).

The group \(S_3\) is solvable (and not Abelian!), a solvable series being \(1 \leq A_3 \leq S_3\). Indeed, \(A_3\) has 3 elements, so it is Abelian; \(A_3\) is normal in \(S_3\), and the factor \(S_3/A_3\) is Abelian, since it has 2 elements.

c) \(S_4\) is solvable. A solvable series is \(1 \leq V \leq A_4 \leq S_4\), where \(V = \{1, (12)(34), (13)(24), (14)(23)\}\)\(^{91}\). Check the details!

d) Any nonabelian *simple* group (i.e. having no proper normal subgroups) is not solvable.

Before studying solvable groups, let us recall some elementary results in group theory.

**6.3 Theorem.** (The fundamental isomorphism theorem) Let \(\varphi: G \rightarrow G'\) be a group homomorphism. Then there exists a canonical isomorphism

\[
G/\text{Ker}\varphi \cong \text{Im}\varphi
\]

\[
x\text{Ker}\varphi \mapsto \varphi(x)
\]

---

\(^{90}\) Some authors use in this case the term “subnormal” and call a “normal series” a chain \((S)\) of normal subgroups of \(G\).

\(^{91}\) \(V\) is also called the *Klein group* (the “Viergruppe”).
6.4 **Theorem.** (Second isomorphism theorem) Let \( G \) be a group, let \( H \) and \( N \) be subgroups of \( G \), such that \( N \) is a normal subgroup in the group \((H, N)\) generated by \( H \cup N \) in \( G \). Then \((H, N) = HN = \{hn \mid h \in H, \ n \in N\} = NH\), \( N \cap H \triangleleft H \) and there is a canonical isomorphism

\[
HN/N \cong H/(N \cap H).
\]

If \( H, N \) are subgroups of \( G \) satisfying the conditions in the previous theorem, \( HN \) or \( NH \) denotes the subgroup generated by \( H \cup N \).

6.5 **Theorem.** (Third isomorphism theorem) Let \( G \) be a group, let \( A \) and \( B \) be normal subgroups of \( G \) such that \( A \leq B \). Then \( B/A \triangleleft G/A \) and there is a canonical isomorphism

\[
\frac{G/A}{B/A} \cong \frac{G}{B}.
\]

6.6 **Proposition.** (The modularity property) Let \( G \) be a group and let \( A, B, N \) be subgroups of \( G \), such that \( N \) is normal and \( B \leq A \). Then \( A \cap (BN) = B \cap (AN) \).

6.7 **Proposition.** Let \( G \) be a group.

a) If \( G \) is solvable, then each subgroup of \( G \) is solvable.

b) If \( G \) is solvable, then each factor group of \( G \) is solvable.

c) If \( H \triangleleft G \) and \( H \) and \( G/H \) are solvable, then \( G \) is solvable. In other words, if

\[
1 \to H \to G \to F \to 1,
\]

is an exact sequence\(^{92}\) of groups and group homomorphisms, then \( G \) is solvable if and only if \( H \) and \( F \) are solvable.

\(^{92}\) The notion of exact sequence of groups and group homomorphisms is defined exactly as in the case of modules.
Proof. \( a \) Let \( 1 = G_0 \leq G_1 \leq \ldots \leq G_n = G \) be a solvable series of \( G \), let \( H \leq G \) and let \( H_i := G_i \cap H \). We claim that \( 1 = H_0 \leq H_1 \leq \ldots \leq H_n = H \) is solvable series of \( H \).

Fix \( i, 1 \leq i \leq n \). The canonical homomorphism \( \varphi : H_i \to G_i/G_{i-1} \) \( (\varphi(x) = xG_{i-1}, \ \forall x \in H_i) \) has kernel \( H_{i-1} \). Thus \( H_{i-1} \triangleleft H_i \) and \( H_i/H_{i-1} \cong \text{Im} \varphi \leq G_i/G_{i-1} \), which is Abelian.

\( b \) Let \( F \) be a factor group of \( G \). There exists a surjective homomorphism \( \varphi : G \to F \). If \( 1 = G_0 \leq G_1 \leq \ldots \leq G_n = G \) is a solvable series of \( G \), let \( F_i := \varphi(G_i) \). We claim that \( 1 = F_0 \leq F_1 \leq \ldots \leq F_n = F \) is a solvable series of \( F \). Since \( G_i-1 \triangleleft G_i \) and \( \varphi \) is surjective, \( F_{i-1} \triangleleft F_i \). Let \( H = \text{Ker} \varphi \). Then \( \text{Ker}(\varphi|_{G_i}) = H \cap G_i \), so

\[
F_i \cong G_i/\text{Ker}(\varphi|_{G_i}) = G_i/(H \cap G_i) \cong (G_i/H)/H.
\]

We used the fundamental and the second isomorphism theorems. Therefore:

\[
\frac{F_i}{F_{i-1}} \cong \frac{G_iH/H}{G_{i-1}H/H} = \frac{G_iH}{G_{i-1}H} = \frac{G_i/(G_{i-1}H)}{G_{i-1}H} = \frac{G_i}{G_{i-1}H} \cong \frac{G_i}{G_{i-1}(H \cap G_i)}.
\]

We used successively: the third, the second isomorphism theorems and the modularity property for \( H \triangleleft G \) and \( G_{i-1} \triangleleft G_i \). But:

\[
\frac{G_i}{G_{i-1}(H \cap G_i)} \cong \frac{G_i/G_{i-1}}{G_{i-1}(H \cap G_i)/G_{i-1}}.
\]

The last group is Abelian (it is a factor group of \( G_i/G_{i-1} \)).

\( c \) Any subgroup of \( G/H \) is of the form \( G'/H \), where \( G' \leq G \), with \( H \leq G \). Therefore, a solvable series for \( G/H \) is of the form \( 1 = G_0/H \leq G_1/H \leq \ldots \leq G_n/H = G/H \), where

\[
H = G_0 \leq G_1 \leq \ldots \leq G_n = G.
\]

If \( 1 = H_0 \leq H_1 \leq \ldots \leq H_m = H \) is a solvable series for \( H \), then \( 1 = H_0 \leq H_1 \leq \ldots \leq H_m = G_0 \leq G_1 \leq \ldots \leq G_n = G \) is a solvable series for \( G \). \( \square \)
6.8 Proposition. Let $G$ be a group. If $G$ has a normal series whose factors are solvable groups, then $G$ is solvable.

Proof. If the series has length 2, there exists $H < G$ such that $H$ and $G/H$ are solvable and the previous result applies. Continue by induction on the length of the series.

The finite solvable groups have the following property, which is essential in the theory of solvability by radicals.

6.9 Proposition. A finite group is solvable if and only if it has a normal sequence whose factors are cyclic groups of prime orders.

Proof. Of course, a group with the property above is solvable.

Conversely, let $G$ be finite solvable. If $|G| \leq 3$, all is clear. We suppose the statement is true for any solvable group whose order is less than $|G|$ and we prove for $G$.

If $G$ is Abelian, then use the following lemma.

Assume now that $G$ is not Abelian. The definition of solvability implies the existence of a proper normal subgroup $H$ of $G$ such that $H$ is solvable and $G/H$ is Abelian. Since $H$ and $G/H$ have orders less than $|G|$, the induction hypothesis says that $H$ and $G/H$ have each a normal series with cyclic factors of prime orders. As in the proof of prop. 6.7.c), gluing these series yields a solvable series of $G$, with cyclic factors of prime orders.

6.10 Lemma. A finite Abelian group has a normal sequence whose factors are cyclic groups of prime orders.

Proof. This is an easy consequence of the structure theorem of the Abelian finite groups. We give a proof that does not use this result.

Assume first that $(G, \cdot)$ is cyclic of order $n$, generated by $x \in G$. If $n$ is prime, we are finished; if not, for any $d$ dividing $n$, there exists a subgroup of order $d$ (generated by $x^{n/d}$). Thus, taking a prime divisor $p$ of $n$, there exists a subgroup $H$ of order $p$. The factor group $G/H$ and $H$ have orders less than $|G|$. Apply now an induction, as above.
If $G$ is not cyclic, let $x \in G$, $x \neq 1$. Thus, the cyclic subgroup $C$ generated by $x$ is not equal to $G$. Apply an induction for $C$ and $G/C$, of orders less than $|G|$. 

6.11 Proposition. Let $G$ be a finite group, whose order is a power of a prime $p$ (such a group is called a $p$-group). Then $G$ is solvable. In particular, there exists a normal series of $G$ whose factors are cyclic groups of order $p$.

Proof. Let $|G| = p^n$. If $n = 1$, then $G$ is cyclic, and solvable. If $n > 1$, then the center of $G$, $C(G) \neq 1$. Indeed, in the conjugacy classes formula (IV.3.15.e),

$$|G| = |C(G)| + \sum_{a \in S} [G : C(a)],$$

(where $S$ is a system of representatives of the conjugacy classes of $G$ with at least 2 elements), $|G|$ and $[G : C(a)]$ are powers of $p$, so $|C(G)|$ is a multiple of $p$ (and is nonzero, since $1 \in C(G)$).

Consequently, $1 \neq C(G) \triangleleft G$ and the factor group $G/C(G)$ is still a $p$-group, whose order is less than $|G|$. Using an induction argument, we obtain that $G/C(G)$ is solvable. Since $C(G)$ is Abelian, thus solvable, we obtain that $G$ is solvable. 

The next proposition says that among the symmetric groups $S_n$, the solvable groups are exactly those in Example 6.2. The proof is not included, being often encountered in introductory Algebra texts (see for instance HUNGERFORD [1974])

6.12 Proposition. a) If $n \leq 4$, then $S_n$ is solvable.

b) If $n \geq 5$, then $S_n$ is not solvable. More precisely, the alternating subgroup $A_n$ is simple (has no proper normal subgroups) and noncommutative, so it is not solvable.
Index

A
adjoining elements to a field.............. 175
algebra ........................................ 334
algebra homomorphism .............. 335
algebraic (element) ..................... 177
algebraic closure (absolute) ....... 198
algebraic closure (relative) ..... 189
algebraic generators .................... 226
algebraic integer ....................... 27
algebraic number ....................... 178
algebraically independent (set) 223
alternating group ....................... 374
annihilator .............................. 104, 133
arithmetically equivalent matrices 121
arrow ........................................ 364
Artin-Schreier theorem ............ 302
ascending chain condition ........ 34
associated in divisibility ........ 12
autodual .................................... 370

B
basis change matrix ................. 113
basis of a module ................. 106
bilinear function ................... 307
nondegenerate ....................... 307
bimorphism ....................... 367

C
canonical homomorphism ............ 75
canonical injections ................. 86
canonical projections ............. 81
canonical surjection .................. 75
category ................................ 364
canonical surjection ............... 75
category ................................ 364
canonical surjection ............... 75
category ................................ 364
canonical surjection ............... 75
category ................................ 364
Chinese remainder theorem .......... 141
Cayley-Hamilton theorem .......... 163
center of a ring ....................... 334
characteristic exponent ............ 173
characteristic matrix .............. 290
characteristic of a ring .......... 172
characteristic polynomial ........ 290
class of a matrix ...................... 161
class of an endomorphism .......... 161
Chevalley theorem .................. 223
Chevalley theorem .................. 223
Chinese remainder theorem ....... 141
class ............................................ 363
codomain .................................... 364
coefficient of a polynomial......... 347
cokernel ........................................ 96
comaximal ideals..................... 141
complement ................................. 90
complex of modules ..................... 96
composite of fields ..................... 175
composition of morphisms ......... 365
congruent modulo n...................... 26
conjugate elements ................. 235
conjugate extension ..................... 240
constant term ............................. 347
constructible complex number .. 278
constructible point ..................... 276
constructible real number ......... 277
content of a polynomial............... 41
coordinates of an element in a basis
............................................... 106
coprime ......................................... 14
coproduct ..................................... 86
cyclotomic extension ..................... 213
cyclotomic polynomial ..................... 214

determinant of an endomorphism \[162\]
diagonal matrix ........................... 121
direct product
   of homomorphisms ................... 92
   of modules .............................. 81
   of modules .............................. 83
direct sum
   of modules .............................. 85
   of submodules ......................... 90
direct summand ......................... 90
discriminant
   of a polynomial ...................... 320
   of an element ......................... 320
divisibility .................................... 10
division with remainder theorem .. 22
divisor ........................................ 11
domain ....................................... 11, 364
dual of a group ........................... 306
duality ....................................... 369

E
eigenvalue .............................. 165
eigenvector .............................. 165
Eisenstein criterion ....................... 47
element
   integral ................................... 52
elementary divisors ................... 147
elementary divisors (of a matrix) ... 154
elementary divisors (of an
   endomorphism) ......................... 154
elementary transformations ......... 123
endomorphism ............................. 68
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyclic</td>
<td>153</td>
</tr>
<tr>
<td>endomorphism ring</td>
<td>61</td>
</tr>
<tr>
<td>epimorphism</td>
<td>76, 367</td>
</tr>
<tr>
<td>Euclidian Algorithm</td>
<td>23</td>
</tr>
<tr>
<td>Euclidian domain</td>
<td>22</td>
</tr>
<tr>
<td>Euler phi function</td>
<td>213</td>
</tr>
<tr>
<td>evaluation homomorphism</td>
<td>346</td>
</tr>
<tr>
<td>exponent (of a group)</td>
<td>303</td>
</tr>
<tr>
<td>exponent of a group</td>
<td>221</td>
</tr>
<tr>
<td>extension</td>
<td>170</td>
</tr>
<tr>
<td>Abelian</td>
<td>299</td>
</tr>
<tr>
<td>algebraic</td>
<td>187</td>
</tr>
<tr>
<td>cyclic</td>
<td>299</td>
</tr>
<tr>
<td>finite</td>
<td>172</td>
</tr>
<tr>
<td>finitely generated</td>
<td>175</td>
</tr>
<tr>
<td>Galois</td>
<td>267</td>
</tr>
<tr>
<td>Kummer</td>
<td>303</td>
</tr>
<tr>
<td>n-radical</td>
<td>314</td>
</tr>
<tr>
<td>purely transcendental</td>
<td>224</td>
</tr>
<tr>
<td>radical</td>
<td>314</td>
</tr>
<tr>
<td>simple</td>
<td>175</td>
</tr>
<tr>
<td>transcendental</td>
<td>188</td>
</tr>
<tr>
<td>extension of a module</td>
<td>97</td>
</tr>
<tr>
<td>external direct sum of modules</td>
<td>87</td>
</tr>
<tr>
<td>$F$</td>
<td></td>
</tr>
<tr>
<td>factor algebra</td>
<td>336</td>
</tr>
<tr>
<td>factorization algorithm</td>
<td>48</td>
</tr>
<tr>
<td>field</td>
<td>169</td>
</tr>
<tr>
<td>algebraically closed</td>
<td>198</td>
</tr>
<tr>
<td>of rational fractions</td>
<td>348</td>
</tr>
<tr>
<td>field extension</td>
<td>170</td>
</tr>
<tr>
<td>field generated by a set</td>
<td>174</td>
</tr>
<tr>
<td>fixed field</td>
<td>233</td>
</tr>
<tr>
<td>fixed subfield</td>
<td>172</td>
</tr>
<tr>
<td>formal derivative</td>
<td>195</td>
</tr>
<tr>
<td>$\mathbb{F}_p$</td>
<td>171</td>
</tr>
<tr>
<td>fraction</td>
<td>356</td>
</tr>
<tr>
<td>free resolution</td>
<td>109</td>
</tr>
<tr>
<td>free subset of a module</td>
<td>105</td>
</tr>
<tr>
<td>Frobenius endomorphism</td>
<td>210</td>
</tr>
<tr>
<td>Frobenius theorem</td>
<td>163</td>
</tr>
<tr>
<td>functor</td>
<td></td>
</tr>
<tr>
<td>contravariant</td>
<td>371</td>
</tr>
<tr>
<td>covariant</td>
<td>370</td>
</tr>
<tr>
<td>fidel</td>
<td>371</td>
</tr>
<tr>
<td>forgetful</td>
<td>372</td>
</tr>
<tr>
<td>fully faithful</td>
<td>371</td>
</tr>
<tr>
<td>plin</td>
<td>371</td>
</tr>
<tr>
<td>functor morphism</td>
<td>372</td>
</tr>
<tr>
<td>fundamental isomorphism theorem</td>
<td>77</td>
</tr>
<tr>
<td>Fundamental Theorem of Algebra</td>
<td>205</td>
</tr>
<tr>
<td>fundamental theorem of Galois</td>
<td></td>
</tr>
<tr>
<td>Galois connections</td>
<td>233</td>
</tr>
<tr>
<td>Galois Field</td>
<td>211</td>
</tr>
<tr>
<td>Galois group</td>
<td></td>
</tr>
<tr>
<td>of a field extension</td>
<td>233</td>
</tr>
<tr>
<td>of a polynomial</td>
<td>233</td>
</tr>
<tr>
<td>Galois group of a polynomial</td>
<td>314</td>
</tr>
</tbody>
</table>
Gauss integers .................................. 12
GCD ............................................. 14
GCD-domain ................................ 15
general equation of degree $n$...... 318
generic polynomial of degree $n$ .. 318
$GL(n, R)$...................................... 114
Gr$366$  
greatest common divisor .......... 14
group  
  solvable................................. 373
group of units ............................ 13

$H$
Hilbert Satz 90......................... 300
  additive version..................... 301
homomorphism
  of module extensions ............ 97

$I$
ideal........................................ 20, 331
  maximal ............................... 332
  prime................................. 332
  proper................................... 331
idempotent.................................. 91
identity morphism.................... 365
image........................................ 71
independent family of submodules90
indeterminate............................ 342
initially constructible point...... 275
inseparable
  element .................................. 244
  extension................................. 245
inseparable degree................... 258
integer division theorem .......... 44
integers modulo $n$ ................. 26
integral domain .................. 11
intermediate field .................. 171
  proper................................... 171
internal direct sum ................. 90
invariant factors ................... 137
invariant factors (of a matrix) ... 154
invariant factors (of an endomorphism) .... 154
Invariant factors theorem ....... 131
invariant subspace............... 152
inverse (of a morphism)........... 367
$Irr(x, K)$................................. 179
irreducible............................. 18
isomorphism.............................. 367

$J$
Jordan canonical form of a matrix164
Jordan canonical matrix .......... 157
Jordan cell.............................. 156

$K$
kernel ...................................... 71
$K$-homomorphism .................. 171
Kummer Theory......................... 303
  additive................................. 311
  multiplicative ...................... 308

$L$
Lagrange interpolation polynomial54
lattice ................................. 176
LCM ........................................ 14
leading coefficient .......... 347
leading monomial ............. 351
least common multiple ....... 14
lexicographic order .......... 350
linear combination .......... 63
linearly dependent .......... 104, 105
linearly independent ....... 104, 105
Lüroth’s theorem .......... 187

\textbf{M}

matrix
  canonic ally diagonal ....... 121
elementary ....................... 122
in Smith normal form ...... 121
matrix companion of a polynomial
 ......................................... 156
matrix of a homomorphism ... 112
minimal polynomial .......... 179
minimal polynomial (of a matrix) 154
minimal polynomial (of an endomorphism) ..... 154
\textit{Mod}-\textit{R} ......................... 366
module .................................... 59
decomposable ..................... 140
factor .................................... 75
finitely generated ............ 65
free ....................................... 106
indecomposable ................ 140
left .................................... 59
of fractions ........................ 360
right .................................... 59
simple ................................... 80
module automorphism ........... 71
module homomorphism .......... 68
module isomorphism ............. 71
monic polynomial ............... 347
monoid .................................. 336
monomial .............................. 342
monomorphism .................. 76, 367
morphism .............................. 364
multiple .............................. 11
multiple root ..................... 45
multiplicative set ............. 355
  saturated .......................... 360
multiplicatively closed set ... 355
multiplicity of a root .......... 45, 194
mutually prime ................... 14

\textbf{N}

natural isomorphism ........... 373
natural transformation ........ 372
Newton's identities ............ 353
Noetherian ring ................... 35
norm ................................... 27, 291
normal closure .................... 241
normal extension ............... 238
normal series ...................... 374
numerator ......................... 356

\textbf{O}

object
  final .................................. 367
free......................................... 372
initial.................................. 367
zero ...................................... 367
object (in a category) .............. 364
opposite of a ring ..................... 60
order ...................................... 130

$P$

$p$-basis.................................. 261
$p$-dimension ............................ 261
perfect field ............................... 246
$p$-group .................................. 378
PID .......................................... 31
polynomial
   elementary symmetric .............. 349
   fundamental symmetric ............ 349
   homogeneous .......................... 347
   monic .................................. 27
   symmetric .............................. 348
polynomial algebras .................... 341
polynomial function ..................... 54
prime ...................................... 18
prime field ................................ 172
primitive element ....................... 175
primitive polynomial ................. 41
principal ideal ......................... 20
principal ideal domain ............... 31
projective linear group .............. 272
proper divisor .......................... 13
$p$-submodule ............................ 133
purely inseparable ..................... 254
purely inseparable closure .......... 255

$Q$

quadratic integer ....................... 27
quaternion group ....................... 208
quaternions, skew field of ........ 207
quotient .................................. 22, 356

$R$

rank of a free module .................. 111
rational integers ...................... 27
remainder ............................... 22
resultant ............................... 324
ring
   integrally closed .................... 52
   of quadratic integers ............... 28
$R$-$Mod$ .................................. 366
root
   multiple .............................. 194
   simple .................................. 194
root of a polynomial .................. 177
root of unity ........................... 212
   primitive ............................. 213

$S$

separable
   element ............................... 244
   extension ............................ 245
   polynomial .......................... 244
separable closure ..................... 253
separable degree ...................... 248
sequence ............................... 97
sequence of modules .................. 96
   exact .................................. 96
semiexact ................................. 96
Set ........................................ 365
similar endomorphisms .......... 152
similar matrices ................. 152
simple root ......................... 45
sink ..................................... 364
smallest element ................. 352
solvable by radicals ............ 314
solvable series ..................... 374
source .................................. 364
split epimorphism ............... 100
split monomorphism ............ 100
split short exact sequence ...... 100
splits .................................... 183
splitting field .................... 201
squarefree .......................... 11
structural homomorphism (of an
   algebra) .............................. 335
subalgebra ............................ 335
   generated by a set................ 335
subcategory ....................... 367
   full .................................. 367
subextension ....................... 171
subfield ................................ 170
   proper ................................ 170
submodule ........................... 62
   cyclic ................................ 65
   generated by a set............... 64
maximal .............................. 67
minimal ................................ 67
   proper ................................ 63
subring generated by a set....... 175
sum of ideals ....................... 20, 331
support ............................. 86, 337
symmetric group .................. 374
system of generators .......... 64

\textbf{T}

term ..................................... 342
torsion module..................... 133
torsion submodule ................. 133
torsion-free module ............. 133
total degree ....................... 347
trace ..................................... 27, 291
trace of a matrix ................. 161
trace of an endomorphism ....... 162
transcendence basis .............. 226
transcendence degree .......... 229
transcendental (element) ..... 177
transcendental number .......... 178
trigonable matrix ............... 167

\textbf{U}

UFD .............................................. 36
\textit{u}-indecomposable space ...... 152
\textit{u}-invariant subspace ........ 152
unique factorization domain ....... 36
unit ........................................ 13
universality property
   of the direct product .......... 82, 83
   of the direct sum ................. 85
   of the free module .............. 107
   of the monoid algebra .......... 343
of the polynomial algebra ...... 344  
of the ring of fractions .......... 358  
$U(R)$ .............................................. 13  

$W$  

Warning Theorem .......................223  
Wedderburn's Theorem..............219  
well ordered .................................352
Bibliography

26. Lang, S. [1964], Algebraic numbers, Addison-Wesley, Reading Massachusetts.
6. Solvable Groups


49. VAN DER WAERDEN, B.L. [1967], *Algebra II* (Fünfte auflage der Modernen Algebra) Springer-Verlag, Berlin.

50. VAN DER WAERDEN, B.L. [1971], *Algebra I* (Achte auflage der Modernen Algebra) Springer-Verlag, Berlin.


