Advanced Algebraic Structures

Groningen, 3rd year bachelor mathematics, 2017
(partly a translation of parts of older lecture notes in Dutch by
L.N.M. van Geemen, H.W. Lenstra, F. Oort, and J. Top.)

J. Top

Contents

preface

These lecture notes contain a translation into English of a number of chapters from the Dutch lecture notes Algebra II (Algebraic Structures) as they were used in the mathematics curriculum of Groningen University during the period 1993–2013. The original Dutch text may be found at `http://www.math.rug.nl/~top/dic.pdf`.

Both the present text and the original are loosely based on another Dutch text on Rings and Fields, called *Algebra II*, written in the late 1970's at the university of Amsterdam by Prof.dr. F. Oort and Prof.dr. H.W. Lenstra. In the 80's L.N.M. van Geemen at Utrecht university added some chapters to the text, and in the 90's in Groningen I included a number of changes.

The present translation consists of two parts. The first one (*Algebraic Structures*) deals with basic concepts and properties of rings and fields as presented in the chapters $1-5$, $7-9$, and 12 of the Dutch notes. The second one (*Advanced Algebraic Structures*) discusses the chapters 8 (automorphisms and splitting fields, in slightly more detail than originally in order to facilitate treating Galois theory), a discussion of basic Galois theory based on previous lecture notes by Prof. M. van der Put and me, the chapters 11 (symmetric polynomials) and 10 (algebraically closed fields) as well as an extended version of chapter 14 (cyclotomic fields), and finally the chapters $6, 13$ on (projective) modules. A short introduction to tensor products is added.

Groningen, January 2017
Jaap Top

I AUTOMORPHISMS OF FIELDS AND SPLITTING FIELDS

I.1 Basic definitions

We repeat some basic definitions and properties of fields. In fact some of this was already used in the previous chapter.

A field K has a smallest subfield, called the prime field of K. This prime field is either $\mathbb{Q}$, in which case the characteristic of K is 0, or $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for some prime number p, in which case the characteristic of K is p.

Let $K \subset L$ be an extension of fields. Then L can be seen as a vector space over K. The dimension of this vector space is denoted by $[L:K]$. In general $[L:K]$ is infinite. If $[L:K]$ is finite then L is called a finite extension of K of degree $[L:K]$.

Let A be a subset of L, then $K(A)$ denotes the smallest subfield of L containing both A and K. Similarly, $K[A]$ denotes the smallest subring of L containing A and K. For $A = \{a_1,\ldots,a_s\}$ one writes $K(A) = K(a_1,\ldots,a_s)$ and $K[A] = K[a_1,\ldots,a_s]$. An element $a \in L$ is called *algebraic over* K if there is a non-zero polynomial $f \in K[x]$ with $f(a) = 0$. For algebraic a there is a polynomial $F \neq 0$ of minimal degree such that $F(a) = 0$. If F is normalized to be monic, then F is unique and is called the minimal polynomial of a over K. Let F have degree n, then $K(a) = K[a] \cong K[x]/(F)$ is a vector space over K with dimension n and with basis $1, a, \ldots, a^{n-1}$.

An element $a \in L$ which is not algebraic over K is called *transcendental* over K. The obvious ring homomorphism $K[x] \to K[a]$ (i.e., $\sum c_i x^i \mapsto \sum c_i a^i$) is an isomorphism. Thus $K[a]$ is not a field. The field of fractions $K(a)$ of $K[a]$ is in this case isomorphic to $K(x)$, i.e., to the field of rational functions over K.

The field extension $K \subset L$ is called finitely generated if there are elements $a_1,\ldots,a_s \in L$ such that $L = K(a_1,\ldots,a_s)$. The elements $a_1,\ldots,a_s$ are called *algebraically dependent over* K if there is a non-zero polynomial $f \in K[x_1,\ldots,x_s]$ with $f(a_1,\ldots,a_s) = 0$. If such a polynomial f does not exists, then $a_1,\ldots,a_s$ are called *algebraically independent over* K. In the latter case, the obvious homomorphism $K[x_1,\ldots,x_s] \to K[a_1,\ldots,a_s]$ is an isomorphism.Then $K \subset L$ is called a *purely transcendental extension of* K *of transcendence degree* s and L is isomorphic to the field of fractions of the polynomial ring $K[x_1,\ldots,x_s]$. This field of fractions is denoted by $K(x_1,\ldots,x_s)$. It is an exercise (see Exercise 3 below) to show that a finitely generated extension L of K has an intermediate field M (i.e., $K \subset M \subset L$) such that $K \subset M$ is purely transcendental and $M \subset L$ is a finite extension. The *transcendence degree of* L *over* K is defined as the transcendence degree of M over K. It is not at all clear that this definition is a valid one. One has to show that it does not depend on the choice of the intermediate purely transcendental extension. This is, for instance, Theorem 25 in Chapter II of the book *Commutative Algebra* (Vol. 1) by O. Zariski and P. Samuel. A different proof which uses theory of "deriva-

tions" and which is valid only in characteristic zero, is to compute the dimension of the M-vector space consisting of all K-linear maps $D : M \to M$ which satisfy $D(m_1 m_2) = m_1 D(m_2) + m_2 D(m_1)$.

I.2 Homomorphisms of fields

This section reviews some basic results on automorphisms of fields. If K and L are fields then $\phi : K \to L$ is called a *homomorphism of fields* (also called 'field homomorphism') if ϕ is a unitary ring homomorphism. In particular it satisfies $\phi(1) = 1$.

A homomorphism of fields has the property

$$\phi(\frac{1}{a}) = \phi(a)^{-1}, \qquad \phi(\frac{a}{b}) = \frac{\phi(a)}{\phi(b)},$$

since $\phi(a) \cdot \phi(a^{-1}) = \phi(a \cdot a^{-1}) = \phi(1) = 1$ implies $\phi(a^{-1}) = (\phi(a))^{-1}$.

Using $\phi(1 + 1 + \ldots + 1) = 1 + 1 + \ldots + 1$ it follows that the image of the prime field K_0 of K is the prime field L_0 of L. Hence ϕ yields an isomorphism from K_0 to L_0. In case $K \subset L$ then $K_0 = L_0$ and the restriction of ϕ to K_0 is the identity map.

The image $\phi(K)$ of a field homomorphism $\phi : K \to L$ is also a field. Every field homomorphism is injective, since the only ideals in K are (0) and K and $1 \notin \mathrm{Ker}(\phi)$. A field homomorphism need not be surjective. For example the inclusion $\mathbb{R} \hookrightarrow \mathbb{C}$ is not. Even if $K = L$ a field homomorphism needs not be surjective, as illustrated by Example I.2.2.

The composition of field homomorphisms

$$K \xrightarrow{\phi} L \xrightarrow{\psi} M$$

is also a field homomorphism, as one easily verifies.

An interesting and important field homomorphism exists in case $\mathrm{char}(K) = p$:

I.2.1 Theorem. *Let K be a field such that* $\mathrm{char}(K) = p > 0$. *Put*

$$F : K \longrightarrow K, \qquad F : x \mapsto x^p.$$

Then F is a field homomorphism called the Frobenius *homomorphism.*
In case K is finite, F is even a field automorphism.

Proof. Note that $F(1) = 1$ and $F(ab) = (ab)^p = a^p b^p = F(a)F(b)$ (since K is commutative). It remains to show that $F(a + b) = (a + b)^p$ equals $F(a) + F(b) = a^p + b^p$.

Newton's binomium formula, which holds in any commutative ring, implies:

$$(a + b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k}, \qquad \text{with} \quad \binom{p}{k} = \frac{p!}{k!(p-k)!} \in \mathbb{Z}.$$

The numerator of $\binom{p}{k}$ is divisible by p, and because $0 < k < p$ and p is prime, the denominator is not divisible by p. Hence

$$(a + b)^p = a^p + b^p + p \cdot c$$

for some $c \in K$. As $p = 1 + 1 + \ldots + 1$ ($p\times$) and $\mathrm{char}(K) = p$ it follows that $p = 0 \in K$. We conclude $(a + b)^p = a^p + b^p$.

In case K is finite every injective map from K to itself (so for example F) is surjective as well. Hence F is bijective. The fact that F^{-1} is a field homomorphism as well, is easily verified. This shows Theorem I.2.1. ∎

I.2.2 Example. Let $K = \mathbb{F}_p(T)$ be the field of rational functions (= quotients of polynomials) with coefficients in $\mathbb{F}_p$. Put $f(T) = \frac{a_0+a_1T+\ldots+a_nT^n}{b_0+b_1T+\ldots+b_mT^m} \in \mathbb{F}_p(T)$, then

$$\begin{aligned} F(f(T)) &= F\left(\frac{a_0+a_1T+\ldots+a_nT^n}{b_0+b_1T+\ldots+b_mT^m}\right) \\ &= \frac{F(a_0+a_1T+\ldots+a_nT^n)}{F(b_0+b_1T+\ldots+b_mT^m)} \\ &= \frac{F(a_0)+F(a_1)F(T)+\ldots+F(a_n)F(T^n)}{F(b_0)+F(b_1)F(T)+\ldots+F(b_m)F(T^m)} \\ &= \frac{a_0+a_1T^p+\ldots+a_nT^{pn}}{b_0+b_1T^p+\ldots+b_mT^{pm}} = f(T^p). \end{aligned}$$

Here we used $F(a) = a$ for all $a \in \mathbb{F}_p$, the prime field of $\mathbb{F}_p(T)$. The image of the Frobenius homomorphism F therefore consists of all rational functions in the variable T^p with coefficiënts in $\mathbb{F}_p$. In particular F is not surjective on $\mathbb{F}_p(T)$, for example $T \notin \text{image}(F)$ (verify for yourself!). ∎

I.2.3 Definition. If L and L' are extensions of a field K then a *K-homomorphism* $L \to L'$ is a field homomorphism

$$\phi : L \to L' \qquad \text{such that} \quad \phi_{|K} = \mathrm{id}_K.$$

A *K-isomorphism* is a bijective K-homomorphism.
The fields L and L' are called *K-isomorphic* (notation: $L \cong_K L'$) if a K-isomorphism $L \to L'$ exists.
A *K-automorphism* is a K-isomorphism with $L = L'$.

If the fields K and L have the same prime field K_0, then every field homomorphism $K \to L$ is a K_0-homomorphism.

I.2.4 Definition. Given an extension L of a field K we write $\mathrm{Aut}_K(L)$ for the group of all K-automorphisms of L. Here the group law is the composition of maps, and the unit element is id_L.

I.2.5 Theorem. *Let L be a finite extension of the field K and $\alpha \in L$ with minimal polynomial $f_K^\alpha \in K[X]$.*

(i) For every $\phi \in \mathrm{Aut}_K(L)$ we have that $\phi(\alpha)$ is a zero of f_K^α.
(ii) If m is the number of zeros of f_K^α in the field $K[\alpha] \subset L$, then $\#\mathrm{Aut}_K(K[\alpha]) = m$. Here $\#S$ the number of elements of a set S.
(iii) The number of K-automorphisms of $K[\alpha]$ is $\leq \deg(f_K^\alpha)$.

Proof. (i): Write $f_K^\alpha = X^n + a_{n-1}X^{n-1}\ldots+a_1X+a_0$ with $a_i \in K$. Applying $\phi \in Aut_K(L)$ to the equality $0 = f_K^\alpha(\alpha)$ yields:

$$\begin{aligned} 0 &= \phi(\alpha^n + a_{n-1}\alpha^{n-1} + \ldots a_1\alpha + a_0) \\ &= \phi(\alpha)^n + \phi(a_{n-1})\phi(\alpha)^{n-1}\ldots\phi(a_1)\phi(\alpha) + \phi(a_0) \\ &= \phi(\alpha)^n + a_{n-1}\phi(\alpha)^{n-1} + \ldots + a_1\phi(\alpha) + a_0 \\ &= f_K^\alpha(\phi(\alpha)), \end{aligned}$$

where we used $\phi(a_i) = a_i$ (note $a_i \in K$ and $\phi_{|K} = \mathrm{id}_K$). So $\phi(\alpha)$ is a zero of f_K^α.
(ii): Let $\{\alpha_1, \ldots, \alpha_m\} \subset K[\alpha]$ be the zeros of f_K^α in $K[\alpha]$, with $\alpha_1 = \alpha$. We claim that the map

$$\Delta : \mathrm{Aut}_K(K[\alpha]) \longrightarrow \{\alpha_1, \alpha_2, \ldots, \alpha_m\}, \qquad \Delta(\phi) := \phi(\alpha)$$

is a bijection.

From (i) we know that Δ is well defined. We now show that Δ is injective. Every $x \in K[\alpha]$ can be given in a unique way as $x = x_0 + x_1\alpha + \ldots + x_{n-1}\alpha^{n-1}$ with $x_i \in K$. Then

$$\begin{aligned} \phi(x) &= \phi(x_0) + \phi(x_1)\phi(\alpha) + \ldots + \phi(x_{n-1})\phi(\alpha)^{n-1} \\ &= x_0 + x_1\phi(\alpha) + \ldots + x_{n-1}\phi(\alpha)^{n-1}, \end{aligned}$$

so $\phi(x)$ is completely determined by $\phi(\alpha) = \Delta(\phi)$. Therefore, if $\Delta(\phi) = \Delta(\psi)$, then $\phi(x) = \psi(x)$ for all $x \in K[\alpha]$, so $\phi = \psi$.

In order to prove surjectivity, given any zero $\beta \in K[\alpha]$ of f_K^α we have to construct a K-automorphism ϕ with $\phi(\alpha) = \beta$. This is done as follows. Take

$$\mathrm{ev}_\beta : K[X] \longrightarrow K[\alpha]$$

the evaluation homomorphism in β. Since β is a zero of $f_K^\alpha \in K[X]$ it follows that $f_K^\alpha \in \mathrm{Ker}(\Phi_\beta)$. However f_K^α is irreducible in $K[X]$, hence $\mathrm{Ker}(\mathrm{ev}_\beta) = (f_K^\alpha)$. So ev_β induces an injective K-homomorphism $\overline{\mathrm{ev}_\beta} : K[X]/(f_K^\alpha) \to K[\alpha]$ with image $K[\beta] \subset K[\alpha]$. The first isomorphism theorem for rings now implies $K[\beta] \cong K[X]/(f_K^\alpha)$; the given isomorphism is K-linear, hence $K[\beta]$ is a linear subspace of $K[\alpha]$ with dimension $\dim_K(K[X]/(f_K^\alpha)) = \dim_K(K[\alpha])$. This dimension is finite and therefore $K[\beta] = K[\alpha]$; in other words $\overline{\mathrm{ev}_\beta}$ is an isomorphism. Analogous to the above, write $\overline{\mathrm{ev}_\alpha}$ for the isomorphism $K[X]/(f_K^\alpha) \xrightarrow{\cong} K[\alpha]$ induced by the evaluation homomorphism ev_α. We have isomorphisms:

$$\begin{array}{ccc} & & K[\alpha] \\ & \overset{\overline{\mathrm{ev}_\alpha}}{\nearrow} & \\ K[X]/(f_K^\alpha) & & \\ & \underset{\overline{\mathrm{ev}_\beta}}{\searrow} & \\ & & K[\beta] = K[\alpha] \end{array}$$

and hence a K-automorphism $\phi := \overline{\mathrm{ev}_\beta} \circ \overline{\mathrm{ev}_\alpha}^{-1}$: $K[\alpha] \xrightarrow{\cong} K[\alpha]$. The definition of evaluation homomorphisms shows

$$\overline{\mathrm{ev}_\alpha}^{-1}(\alpha) = X + (f_K^\alpha) \qquad \text{and} \quad \overline{\mathrm{ev}_\beta}(X + (f_K^\alpha)) = \beta,$$

hence $\phi(\alpha) = \beta$.
(iii) is immediate from (ii). This finishes the proof. ∎

I.2.6 Example. Let $f \in K[X]$ be a monic irreducible polynomial of degree 2 and put $L = K[X]/(f)$. We write $\alpha := X + (f) \in L$ which is a zero of f in $L = K[\alpha]$. Writing $f = X^2 + aX + b$ one calculates (using long division if necessary) that in $L[X]$:

$$X^2 + aX + b = (X - \alpha)(X - (-a - \alpha)).$$

We now distinguish two cases:

$$(i)\ \ \alpha = -a - \alpha, \qquad (ii)\ \ \alpha \neq -a - \alpha.$$

In the first case $2\alpha = -a$ follows. If $2 \neq 0$ in K then $\alpha = -a/2 \in K$ contradicting the irreducibility of f. Hence, the first case is only possible for $\mathrm{char}(K) = 2$ and $a = 0$. An example of such an irreducible f is $X^2 + T \in \mathbb{F}_2(T)[X]$, a polynomial over the field of rational functions in the variable T with coefficients in $\mathbb{F}_2$. (Were f reducible then $(g/h)^2 = T$ so $g^2 = Th^2$ for certain $g, h \in \mathbb{F}_2[T]$; comparing degrees shows this is impossible.)

In the remaining case (ii) f has two distinct zeros in L, hence $\#\mathrm{Aut}_K(L) = 2$ by Theorem I.2.5. All groups consisting of two elements are isomorphic, so

$$\mathrm{Aut}_K(L) = \{id_L, \phi\} \cong \mathbb{Z}/2\mathbb{Z}.$$

Here $\phi(\alpha) = -a - \alpha$ and $\phi(x + y\alpha) = \phi(x) + \phi(y)\phi(\alpha) = x + y(-a - \alpha)$ for $x, y \in K$).

A well known example is $K = \mathbb{R}$ and $f = X^2 + 1$, here ϕ is complex conjugation.

Another example is given by the finite fields $L = \mathbb{F}_p[X]/(X^2 - d)$ with d a non-square in $\mathbb{F}_p$, in particular we must have $p > 2$ here since every $d \in \mathbb{F}_2$ is (its own) square. Then

$$\phi : L \to L, \qquad \phi : x + y\alpha \mapsto x - y\alpha, \qquad x, y \in \mathbb{F}_p$$

is the nontrivial field automorphism.

On the other hand Theorem I.2.1 yields an automorphism as well, namely the Frobenius automorphism F. We have

$$F(x + y\alpha) = F(x) + F(y)F(\alpha) = x + y\alpha^p$$

since $x, y \in \mathbb{F}_p$, the prime field of L. Claim:

$$F = \phi.$$

As $\#\mathrm{Aut}_{\mathbb{F}_p}(L) = 2$ it suffices to show $F \neq id_L$. In case $F = \mathrm{id}_L$ the polynomial $X^p - X$ would have $\#L = p^2$ zeros in L, a contradiction. Conclusion: $F \neq \mathrm{id}_L$ and therefore $F = \phi$.

From $F = \phi$ one deduces $\alpha^p = F(\alpha) = \phi(\alpha) = -\alpha$. As $p > 2$ is prime and therefore odd, we write $p = 2(\frac{p-1}{2}) + 1$ with $\frac{p-1}{2} \in \mathbb{Z}$. Then

$$-\alpha = \alpha^p = (\alpha^2)^{\frac{p-1}{2}}\alpha = d^{\frac{p-1}{2}}\alpha.$$

This shows $d^{\frac{p-1}{2}} = -1$ in $\mathbb{F}_p \subset L$. ∎

I.3 Solving polynomial equations; splitting fields

Let $f \in K[x]$ be a non-constant polynomial. In general, f does not split as a product of linear factors in $K[x]$ because K does not always contain all the solutions of $f(a) = 0$. We would like to "find" or to "construct" a larger field which contains all the roots of f. This is formalized in the next definition.

I.3.1 Definition. A *splitting field* L of f over K is a field extension such that:
(a) f splits in $L[x]$ as a product of linear factors.
(b) Let $a_1, \ldots, a_s$ denote the zeros of f in L, then $L = K(a_1, \ldots, a_s)$.

I.3.2 Proposition.
(1) *A splitting field exists.*
(2) *Let L_1, L_2 be two splitting fields for f over K. Then there exists a K-linear isomorphism of the fields L_1, L_2.*

Proof. (1) One uses induction with respect to the degree of f. Take a (monic) irreducible factor g (in $K[x]$) of f. Then $K_1 := K[y]/(g(y))$ contains a zero, say α of g. Thus f factors in $K_1[x]$ as $(x - \alpha)h$. By induction, a splitting field L for h over K_1 exists. It is easily seen that L is also a splitting field for f over K. We note, in passing, that $[L : K] < \infty$.
(2) We will make a proof of a somewhat more general statement:
Two fields K_1 and K_2 and an isomorphism $\phi_0 : K_1 \to K_2$ are given. Extend ϕ_0 to a ring isomorphism $K_1[x] \to K_2[x]$ by $\phi_0(\sum a_i x^i) = \sum \phi_0(a_i)x^i$. Suppose $f_1 \in K_1[x]$ and $f_2 \in K_2[x]$ satisfy $\phi_0(f_1) = f_2$. Let L_1, L_2 denote two splitting fields for f_1 and f_2 over

K_1 and K_2, respectively. Then ϕ_0 extends to an isomorphism between the field L_1 and L_2.

We will construct the desired isomorphism $\phi : L_1 \to L_2$ step by step. The lowest level, $\phi_0 : K_1 \to K_2$ is given. Consider an irreducible factor g of f_1 and a zero $\alpha \in L_1$ of g. Then $\phi_0(g)$ is an irreducible factor of $\phi_0(f_1) = f_2$. Hence $\phi_0(g)$ splits completely over L_2. Choose a β in L_2 with $\phi_0(g)(\beta) = 0$. Define $\phi_1 : K_1(\alpha) \to K_2(\beta)$ by the formula $\sum_{i=0}^{n-1} a_i \alpha^i \mapsto \sum_{i=0}^{n-1} \phi_0(a_i)\beta^i$, where all $a_i \in K_1$ and where n is the degree of g. It is easily verified that ϕ_1 is indeed an isomorphism of fields. Now replace K_1, K_2, f_1, f_2 by $K_1(\alpha), K_2(\beta), f_1/(X-\alpha), f_2/(X-\beta)$. The fields L_1, L_2 are splitting fields over $K_1(\alpha)$ and $K_2(\beta)$ for the polynomials $f_1/(X-\alpha)$ and $f_2/(X-\beta)$. Induction finishes the proof. ∎

I.3.3 Notation. For a field K and a nonzero $f \in K[x]$ the splitting field of f over K (which by Proposition I.3.2 is unique up to K-isomorphisms) is denoted Ω_K^f.

A polynomial $f \in K[x] \setminus K$ is called *separable* if the roots of f in any field extension L of K are distinct. It follows from Proposition I.3.2 that it suffices to verify this for a splitting field L of f over K.

I.3.4 Corollary. *Let $L \supset K$ be the splitting field of a separable nonconstant polynomial $f \in K[x]$. The number of K-linear automorphisms of L equals $[L : K]$.*

Proof. In the situation given in the proof of part (2) of Proposition I.3.2, we compute the dimensions and count the number of choices for extensions. We work again in the more general situation. Suppose $\phi_0 : K_1 \to K_2$ is given. The polynomials f_1 and f_2 are by assumption separable. Thus g and $\phi_0(g)$ are separable. It is obvious that $\phi_1 : K_1(\alpha) \to L_2$ should map α to a root β of $\phi_0(g)$. The number of possibilities for β is the degree of $\phi_0(g)$ (which equals the degree of g). Thus for ϕ_1 there are $\deg(g)$ possibilities. One has $[L_1 : K_1] = [L_1 : K_1(\alpha)][K_1(\alpha) : K_1]$. By induction the number of extensions $L_1 \to L_2$ of a given ϕ_1 is equal to $[L_1 : K_1(\alpha)]$. This completes the proof. ∎

I.4 Exercises

1. Let K be a field. Show that the map

$$\phi : K(X) \longrightarrow K(X), \qquad f \mapsto f(X+1)$$

is a field automorphism. Find the order of ϕ in the group $\mathrm{Aut}(K(X))$. (The answer depends on the characteristic $\mathrm{char}(K)\ldots$)

2. Describe a splitting field of $X^2 - 101$ over $\mathbb{Q}$.
3. Let L be a splitting field of f over K and $f = \prod_{i=1}^{n}(X - \alpha_i)$ in $L[X]$. Prove that $L = K(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$ (so α_n is omitted!).
4. Suppose $f \in K[X]$ is monic of degree n. Prove: $[\Omega_K^f : K]$ divides $n!$ (Hint: use the construction in the proof of I.3.2.)
5. Prove that $L = \mathbb{Q}(\sqrt[4]{2}, i)$ is a splitting field of $X^4 - 2$ over $\mathbb{Q}$. Determine $[L : \mathbb{Q}]$ and $\sharp\mathrm{Aut}(L)$.
   Prove that $\mathrm{Aut}_{\mathbb{Q}(i)}(L) \cong \mathbb{Z}/4\mathbb{Z}$.
6. Let $\zeta \in \mathbb{C}$ be a zero of $f = X^4 + X^3 + X^2 + X + 1$. Show that $\zeta^5 = 1$ and that $\zeta^2, \zeta^3, \zeta^4$ are zeros of f as well. Prove that $\mathbb{Q}(\zeta)$ is a splitting field of f over $\mathbb{Q}$. Determine $\mathrm{Aut}(\mathbb{Q}(\zeta))$.
7. Prove that $\Omega_{\mathbb{Q}}^{X^2-2} \ncong \Omega_{\mathbb{Q}}^{X^2-3}$ and that $\Omega_K^{X^2-\bar{2}} \simeq \Omega_K^{X^2-\bar{3}}$ for $K = \mathbb{F}_5$.
8. Prove that $\mathbb{Q}(i, \sqrt{2})$ is a splitting field of $f_{\mathbb{Q}}^{i+\sqrt{2}}$ over $\mathbb{Q}$.
   Show that $\mathrm{Aut}(\mathbb{Q}(i, \sqrt{2})) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.
9. Let L be a splitting field of f over K and put $n = \deg(f)$.
   (a) Prove: any K-automorphism of L permutes the zeros of f in L,
   (b) Prove: the group $\mathrm{Aut}_K(L)$ is isomorphic to a subgroup of S_n;
   (c) Show that $\#\mathrm{Aut}_K(L)$ is a divisor of $n!$.
10. Prove: $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3})) \simeq S_3$.
11. Take $f = X^3 + X^2 - 2X - 1 \in \mathbb{Q}[X]$ and let $\alpha \in \Omega_{\mathbb{Q}}^f$ be a zero of f.
    Compute $f_{\mathbb{Q}}^{\alpha^2-2}$ and show that $\mathbb{Q}(\alpha) = \Omega_{\mathbb{Q}}^f$. Deduce that $\mathrm{Aut}(\Omega_{\mathbb{Q}}^f) \cong \mathbb{Z}/3\mathbb{Z}$.

II GALOIS THEORY

II.1 Galois extensions and examples

For our purposes a pleasant definition of a *Galois extension* $K \subset L$ is:

II.1.1 Definition. $L \supset K$ is called a Galois extension of K if L is a splitting field of a separable polynomial f over K.

II.1.2 Definition. The *Galois group* $\mathrm{Gal}(L/K)$ of the extension $K \subset L$ is the group of all K-linear automorphisms of L.

Observe that indeed $\mathrm{Gal}(L/K)$ is a group, with composition of automorphisms as group law and the identity automorphism as the unit element .

II.1.3 Lemma. *Let $K \subset L$ be a Galois extension and suppose that $a \in L$ is invariant under the action of* $\mathrm{Gal}(L/K)$*. Then $a \in K$.*

Proof. From Corollary I.3.4 we know that $[L : K] = \#\mathrm{Gal}(L/K)$. Let f be a polynomial in $K[x]$ such that L is its splitting field over K. Observe that L is also the splitting field of f over the field $K(a)$. The assumption that $a \in L$ is invariant implies $\mathrm{Gal}(L/K) = \mathrm{Gal}(L/K(a))$. Thus $[L : K] = [L : K(a)]$ and hence $[K(a) : K] = 1$, which means that $a \in K$. ∎

II.1.4 Corollary. *Let $K \subset L$ be a Galois extension and $a \in L$. The minimal polynomial F of a over K is separable and all its roots are in L.*

Proof. Let the orbit $\mathrm{Gal}(L/K)a = \{ga |$ for all $g \in \mathrm{Gal}(L/K)\}$ be $\{a_1, \ldots, a_s\}$. Consider the polynomial $G := (x-a_1)\cdots(x-a_s) = x^s + b_{s-1}x^{s-1} + \cdots + b_1 x + b_0$ in $L[x]$. This polynomial is invariant under the action of $\mathrm{Gal}(L/K)$ and hence Lemma II.1.3 implies that $G \in K[x]$. Clearly F divides G and therefore it has the required properties. ∎

II.1.5 Proposition. *Let $K \subset L$ be a finite extension. The following are equivalent:*
(1) *$K \subset L$ is a Galois extension.*
(2) *For every element $a \in L$, the minimal polynomial $F \in K[x]$ of a has the property that all its roots lie in L and are simple.*

Proof. (1)⇒(2) is the statement of Corollary II.1.4.
(2)⇒(1). Take elements $a_1, \ldots, a_s \in L$ such that $L = K(a_1, \ldots, a_s)$. Let F_i be the minimal polynomial of a_i over K. We may suppose that $F_1, \ldots, F_t$ are the distinct

elements in $\{F_1,\dots,F_s\}$. Put $F = F_1\cdots F_t$. Then each F_i is separable and has all its roots in L. Then also F is separable and all its roots are in L. Moreover the set of the roots of F contains $\{a_1,\dots,a_s\}$. Thus L is the splitting field of F over K. ∎

In several textbooks on Galois theory, part (2) of the above proposition is used as the definition of a Galois extension. To be more precise: a finite extension L/K is called *normal* if for every $a \in L$ the minimal polynomial F of a over K splits in $L[x]$ as a product of linear factors. The extension is called *separable* if for every $a \in L$, the minimal polynomial F of a over K is separable. The extension is called Galois if it is both normal and separable.

II.1.6 Remark. Let $f \in K[x]$ be a separable polynomial and let $L \supset K$ be a splitting field. The roots of f in L are say $\{a_1,\dots,a_n\}$; note that $n = \deg(f)$. Every $\sigma \in \mathrm{Gal}(L/k)$ permutes this set. Thus we find a homomorphism $\mathrm{Gal}(L/K) \to S_n$. This homomorphism is injective. Indeed, if $\sigma(a_i) = a_i$ for all i, then σ is the identity since $L = K(a_1,\dots,a_n)$.

II.1.7 Example. The Galois group of the splitting field of x^4-2 over $\mathbb{Q}$ is in this way isomorphic with the subgroup

$$\{(1),(24),(1234),(12)(34),(13)(24),(13),(1432),(14)(23)\}$$

of S_4. Here we identify $k \in \{1,2,3,4\}$ with the zero $a_k = i^{k-1}\sqrt[4]{2}$ of x^4-2. A permutation sending k to ℓ corresponds to an automorphism sending a_k to a_ℓ. ∎

We now give some more examples and elementary properties.

II.1.8 Example. $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ *is not normal.*
This follows using Proposition II.1.5. Indeed, the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is x^3-2. One can consider $\mathbb{Q}(\sqrt[3]{2})$ as a subfield of $\mathbb{R}$. The other two zeros of x^3-2 in $\mathbb{C}$ are $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$, where $\omega = e^{2\pi i/3}$. They do not lie in $\mathbb{R}$ and hence also not in $\mathbb{Q}(\sqrt[3]{2})$. ∎

II.1.9 Example. *If* $[L:K] = 2$ *and the characteristic of* K *is* $\neq 2$*, then* L/K *is Galois.*
Indeed, choose $\alpha \in L \setminus K$. Its minimal polynomial F has the form $F = x^2 + ax + b$. The polynomial F splits in $L[x]$ as $(x-\alpha)(x-\beta)$ with $\beta+\alpha = -a$. If α would equal β, then one finds the contradiction $\alpha = -\frac{a}{2} \in K$. It follows that L is the splitting field of the separable polynomial F over K. ∎

II.1.10 Example. *If the characteristic of* K *is 0, then every finite* L/K *is separable.*
To see this, choose $a \in L$. Its minimal polynomial $F \in K[x]$ is irreducible. The derivative $F' = \frac{dF}{dx}$ of F is not 0 and thus the g.c.d. of F and F' is 1. By Exercise 4 of Section I.3, one has that F is separable. ∎

II.1.11 Example. *A field* K *of characteristic* $p > 0$ *is called* perfect *if every element of* K *is a* p*th power. Every finite extension* L/K *of a perfect field is separable.*
Proof: choose $a \in L$ with minimal polynomial $F \in K[x]$. If F is not separable then the g.c.d. of F and F' is not 1. Since F is irreducible this implies that $F' = 0$. Thus F contains only $p$the powers of x, i.e., $F = \sum a_n x^{pn}$ with $a_n \in K$. Each a_n is written as b_n^p with $b_n \in K$. Then $F = (\sum b_n x^n)^p$, which contradicts the fact that F is irreducible. ∎

II.1.12 Example. *Every finite field is perfect. The field* $\mathbb{F}_p(t)$ *is not perfect.*
Indeed, the finite fields of characteristic p are the $\mathbb{F}_q$ with q a power of p. The "Frobenius map" $Fr : z \mapsto z^p$ on any field of characteristic $p > 0$ is a homomorphism

of the additive group. The kernel is 0 since $z^p = 0$ implies $z = 0$. By counting, one sees that $Fr : \mathbb{F}_q \to \mathbb{F}_q$ is bijective. And $Fr : \mathbb{F}_p(t) \to \mathbb{F}_p(t)$ is not surjective since t is not in the image. ∎

II.1.13 Example. *Put $K = \mathbb{F}_p(t)$ and let $F = x^p - t$. The splitting field of F is not separable over K.*
Indeed, the derivative of F is 0. We note that the splitting field in question can be identified with $\mathbb{F}_p(x)$. ∎

II.2 Galois correspondence and primitive elements

Let a Galois extension L/K with Galois group G be given. One considers two sets, $\mathcal{M}$, the set of the intermediate fields, i.e., the fields M with $K \subset M \subset L$ and the set $\mathcal{G}$ of the subgroups of G.

There are two maps between those sets, $\alpha : \mathcal{M} \to \mathcal{G}$, defined by $\alpha(M) = Gal(L/M)$ and $\beta : \mathcal{G} \to \mathcal{M}$, defined by $\beta(H) = L^H$, i.e., the subfield of L consisting of the elements which are invariant under the action of H. It is obvious that the two maps reverse inclusions. What is often called the "main theorem of Galois theory" is

II.2.1 Theorem (The Galois correspondence)**.**
(1) *α and β are each others inverse.*
(2) *The subgroup $H \in \mathcal{G}$ is normal if and only if $\beta(H)$ is a normal (or Galois) extension of K.*
(3) *Suppose that H is a normal subgroup of G. Then M/K is a Galois extension with Galois group G/H.*

We will first prove a lemma.

II.2.2 Lemma. *Suppose that the finite extension L/K has the property that there are only finitely many intermediate fields. Then there is an $\alpha \in L$ with $L = K(\alpha)$.*

Proof. For a finite field K the proof is quite easy. L is also a finite field and it is known that the multiplicative group L^* of L is cyclic, i.e., there is an element ξ with $L^* = \{\xi^n | n \in \mathbb{Z}\}$. Clearly $L = K(\xi)$. Suppose now that K is infinite and that $L \neq K$. Let $n \geq 1$ be minimal such that $L = K(a_1, \ldots, a_n)$ for certain elements $a_1, \ldots, a_n$. We have to show that $n = 1$. Suppose $n \geq 2$ and consider for every $\lambda \in K$ the element $b_\lambda := a_1 + \lambda a_2$. The fields $K(b_\lambda)$ are intermediate fields for L/K and thus there are $\lambda_1 \neq \lambda_2$ with $M := K(b_{\lambda_1}) = K(b_{\lambda_2})$. The field M contains $a_1 + \lambda_1 a_2$ and $a_1 + \lambda_2 a_2$. It follows that $M = K(a_1, a_2)$ and that $L = K(b_{\lambda_1}, a_3, \ldots, a_n)$. This contradicts the minimality of n. ∎

Proof. of Theorem II.2.1
(1) $\beta\alpha(M)$ is the field $L^{Gal(L/M)}$. Applying Lemma II.1.3 to the Galois extension L/M implies that $L^{Gal(L/M)} = M$. This implies that α is injective. Since $\mathcal{G}$ is finite, also $\mathcal{M}$ is finite. Take $H \in \mathcal{G}$ and put $M = \beta(H) = L^H$. Clearly $H \subset \alpha\beta(H)$. We have to prove equality. According to Lemma II.2.2 one can write $L = M(a)$ for some $a \in L$. Consider the polynomial $G = \prod_{\sigma \in H}(x - \sigma(a)) \in L[x]$. This polynomial is invariant under the action of H. Therefore its coefficients are also invariant under H and belong to M. Then $G \in M[x]$ and the minimal polynomial of a over M divides G. Thus $[L : M] \leq \#H$ and $\#Gal(L/M) \leq \#H$. Since H is a subgroup of $Gal(L/M)$ one finds the required equality $H = Gal(L/M)$.
(2) The subgroup H corresponds to $M = \beta(H) = L^H$. For any $\sigma \in G$ one has

$$\beta(\sigma H \sigma^{-1}) = \sigma(M).$$

Thus H is normal if and only if $\sigma(M) = M$ for all $\sigma \in G$. Using that L/K is normal and separable, one finds that the latter property of M is equivalent with M/K normal (or also Galois).
(3) H is supposed to be normal and thus $M := \beta(H) = L^H$ satisfies $\sigma(M) = M$ for all $\sigma \in G$. Thus one can define a restriction map $Gal(L/K) \to Gal(M/K)$. The kernel of this homomorphism is H. The map is surjective since $\#(G/H)$ is equal to $[M : K]$. ■

II.2.3 Remark. The proof of II.2.1(2) shows a small fact that deserves to be noted. Namely, if L/K is a finite Galois extension and $\sigma \in G := \mathrm{Gal}(L/K)$, then for every intermediate field M, so $K \subset M \subset L$, also $\sigma(M)$ is an intermediate field. Moreover since σ is a K-linear automorphism, $\dim_K(M) = \dim_K(\sigma(M))$.

If this field M equals L^H for some subgroup $H \subset G$, then L/M is Galois and $H = \mathrm{Gal}(L/M)$, and in particular $\#H = [L : M] = [L : K]/\dim_K(M)$.

The equality $\beta(\sigma H \sigma^{-1}) = \sigma(M)$ states that the subgroup $\sigma H \sigma^{-1}$ via the Galois correspondence is associated to the subfield $\sigma(M)$. In other words, the action of σ on the set $\mathscr{M}$ of intermediate fields yields via the Galois correspondence the action 'conjugate by σ' on the set $\mathscr{G}$ of subgroups.

II.2.4 Corollary (The theorem of the primitive element)**.** *For every finite separable extension L/K there is a cyclic element, i.e., an element $a \in L$ with $L = K(a)$.*

Proof. According to Lemma II.2.2, it suffices to show that there are finitely many intermediate fields for L/K. If we can show that L lies in the splitting field $\tilde{L}$ of a separable polynomial, then there are only finitely many intermediate fields for $\tilde{L}/K$ and then also for L/K. Let L/K be generated by elements $a_1, \ldots, a_s$. The minimal polynomial of a_i over K is denoted by F_i. We may suppose that $F_1, \ldots, F_t$ are the distinct minimal polynomials, then $F := F_1 \cdots F_t$ is separable and every a_i is a zero of F. The splitting field $\tilde{L}$ of F contains L and we are done. ■

II.3 Exercises

1. Let $K \subset L$ be fields and $a \in L$. Show that $K[a]$ is a field if and only if a is algebraic over K.

2. Let $K \subset L$ be fields and $S \subset L$ a non-empty *finite* subset. Show that $K[S]$ is a field if and only if every element of S is algebraic over K.
   Find a counterexample with $S = L$ to show that the condition that S is finite cannot be missed.

3. Let $K \subset L$ be a finitely generated field extension. Prove the existence of an intermediate field M such that $K \subset M$ is purely transcendental (or $K = M$) and $[L : M] < \infty$.
4. Prove that $f \in K[x] \setminus K$ is separable if and only if f and its derivative $f' = \frac{d}{dx} f$ are relatively prime.

5. Let L be a splitting field of f over K. Let n be the degree of f. Prove that $[L : K] \leq n!$. Try to make examples with $K = \mathbb{Q}$ and f of degree 3 with $[L : \mathbb{Q}] = 6$.

6. Let L be the splitting field over $\mathbb{Q}$ of the polynomial $x^3 - 3$. Produce an explicit splitting field $L \subset \mathbb{C}$ and find the ($\mathbb{Q}$-linear) field automorphisms of L.

7. The same question for the polynomial $x^8 - 1$ over $\mathbb{Q}$.

8. Show that a finite extension of a finite field is Galois.

9. Show that $\mathbb{F}_{p^n}$ is the splitting field of the polynomial $x^{p^n} - x$ over $\mathbb{F}_p$. Prove that $\mathbb{F}_{p^n}/\mathbb{F}_p$ is Galois and that its Galois group is $\{1, Fr, \ldots, Fr^{n-1}\}$, where Fr is the Frobenius map on $\mathbb{F}_{p^n}$ defined by $Fr(z) = z^p$.

10. Show that $\mathbb{Q}(\sqrt[4]{3}, i)$ is a Galois extension of $\mathbb{Q}(i)$ and compute the corresponding Galois group.

11. Determine the Galois group of $x^{12} - 2$ over $\mathbb{Q}$.

12. Take $q = p^n$ for some prime p, and let $K \supset \mathbb{F}_q$ and $a \in K$, with the property that the polynomial $X^q - X + a$ does not split completely in $K[X]$. Let α be a zero of this polynomial in some splitting field.
    (a) Show that $K(\alpha) \neq K$.
    (b) Show that the extension $K(\alpha) \supset K$ is Galois.
    (c) Show that every $\sigma \in \mathrm{Gal}(K(\alpha)/K)$ satisfies $\sigma(\alpha) = \alpha + t$ for some $t \in \mathbb{F}_q$.
    (d) Show that $\sigma \mapsto \sigma(\alpha) - \alpha$ defines an injective homomorphism of groups: $\mathrm{Gal}(K(\alpha)/K) \to (\mathbb{F}_q, +, 0)$.
    (e) In the special case that K is itself a finite field, observe that the Galois group of a finite extension of finite fields is cyclic, and deduce that $X^q - X + a$ factors as a product of q/p distinct irreducible polynomials of degree p in $K[X]$.
    (f) Take $K = \mathbb{F}_q(t)$ and $X^q - X + t$. Explain why this polynomial is irreducible in $K[X]$, and determine the Galois group of its splitting field over K.

13. Prove that $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ is a Galois extension of $\mathbb{Q}$. Determine the Galois group, all intermediate subfields and a primitive element.

14. The same questions for $\mathbb{Q}(e^{2\pi i/3}, \sqrt[3]{3})/\mathbb{Q}$.

15. Find all subfields of $\mathbb{Q}(\sqrt[4]{2}, i)$ and a primitive element for each of them. Which of these fields are normal over $\mathbb{Q}$?

16. One of the theorems in the Ph.D. thesis of Amol Sasane (Groningen, 2001, advisor Prof. dr. R. Curtain) states that $\tan(\pi/2001) \notin \mathbb{Q}$. Prove this theorem.

17. (a) Prove that for an odd prime number p, the field $\mathbb{Q}(\zeta_p)$ has a unique subfield K with $[K : \mathbb{Q}] = 2$.
    (b) Find a condition on p such that $K \subset \mathbb{R}$.
    Hint: complex conjugation is an element of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$. The question whether or not K is a real field is the same as the question whether complex conjugation is an element of the subgroup $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/K)$.
    (c) Write $\varepsilon(n) := 1$ if n mod p is a square in $\mathbb{F}_p^*$ and $\varepsilon(n) := -1$ otherwise. Show that $K/\mathbb{Q}$ is generated by the element $\sum_{n=1}^{p-1} \varepsilon(n)\zeta_p^n$.

18. Prove that the regular 7-gon cannot be constructed by ruler and compass. Explanation: The admissible operations with ruler and compass are: drawing a line through two points, drawing a circle with given center and radius, intersecting two lines, interseting a line with a circle and intersecting two circles. Hint: Identify the plane with $\mathbb{C}$. The points $\mathbb{Q} \subset \mathbb{C}$ are given. Prove

that every "constructable" point $a \in \mathbb{C}$ gives rise to a "tower of field extensions" $K_n := \mathbb{Q}(a) \supset K_{n-1} \supset \cdots \supset K_0 = \mathbb{Q}$ with $[K_{i+1} : K_i] = 2$ for every i.

19. *Cyclic extensions.*
Let K be a field and $n > 1$ an integer. Suppose that the characteristic of K is 0 or p with $p \nmid n$ and that K contains all the n^{th} roots of unity. In this exercise we want to prove the following statement:
$E \supset K$ is a Galois extension with a cyclic Galois group of order n if and only if $E = K(\alpha)$ where α is a root of an irreducible polynomial $x^n - a \in K[x]$.
(a) Let $f(x) = x^n - a \in K[x]$ be an irreducible polynomial. Show that f is separable. Show that the splitting field E of $f(x)$ over K is of the form $K(\alpha)$ with α a root of $f(x) = 0$. Furthermore, show that the Galois group of E over K is generated by the map defined by $\alpha \mapsto \zeta\alpha$ where ζ is a primitive n^{th} root of unity.
(b) Let E be a Galois extension of K with a cyclic Galois group of order n. Let σ generate the Galois group.
(i) One considers σ as a K-linear map on E as vector space over K. Prove that every eigenvalue λ of σ satisfies $\lambda^n = 1$ and thus belongs to K.
(ii) Prove that there is a basis of eigenvectors for σ. (Hint: Jordan normal form).
(iii) Prove that every eigenvalue has multiplicity 1. (Hint: if $\sigma e_i = \lambda e_i$ for $i = 1, 2$ and $e_1 \neq 0 \neq e_2$, then $\sigma(\frac{e_1}{e_2}) = \frac{e_1}{e_2}$).
(iv) Prove that there is an $\alpha \in E$ with $\alpha \neq 0$, $\sigma(\alpha) = \zeta\alpha$ and ζ a primitive n^{th} root of unity.
(v) Show that the $\sigma^i(\alpha)$, $i = 0, \ldots, n-1$, are all distinct and that therefore the minimal polynomial of α over K is $x^n - a$ with $a = \alpha^n \in K$.

III ROOTS OF UNITY

III.1 Cyclotomic fields over the rationals

Let $n \geq 1$ be an integer and write $\zeta = \zeta_n = e^{2\pi i/n} \in \mathbb{C}$. The subfield $\mathbb{Q}(\zeta) \subset \mathbb{C}$ is the splitting field of $x^n - 1$ over $\mathbb{Q}$, since all the roots of $x^n - 1$ are ζ^k, $k = 0, 1, \ldots, n-1$. For $n \geq 3$, one calls $\mathbb{Q}(\zeta)$ *the nth cyclotomic field*. Since $x^n - 1 \in \mathbb{Q}[x]$ is separable, $\mathbb{Q}(\zeta) \supset \mathbb{Q}$ is a Galois extension. We will determine the degree and the Galois group of this extension.

The minimal polynomial Φ_n of ζ over $\mathbb{Q}$ is called *the nth cyclotomic polynomial*. An explicit general formula for Φ_n is not available. Still we will need some information on Φ_n in order to calculate the Galois group of $\mathbb{Q}(\zeta)/\mathbb{Q}$. Two tools from standard courses in algebra that we will use are:
(1) **Lemma of Gauss**: Let $f \in \mathbb{Z}[x]$ be monic and let $g, h \in \mathbb{Q}[x]$ be monic, too. Then $f = gh$ implies that $g, h \in \mathbb{Z}[x]$.
(2) **Eisenstein's criterion**: Let $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ be such that all its coeffcients, with the exception of a_n, are divisible by a prime number p and p^2 does not divide a_0. Then f is irreducible in $\mathbb{Z}[x]$ and in $\mathbb{Q}[x]$.
The proof of both statements uses "reduction modulo p", i.e., the ring homomorphism $\mathbb{Z}[x] \to \mathbb{F}_p[x]$, given by $f = \sum a_n x^n \mapsto \overline{f} := \sum \overline{a}_n x^n$, where for $a \in \mathbb{Z}$ one has written $\overline{a}$ for its image in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Reduction modulo p will also be used in the proof of the next proposition, as well as the identity $\overline{f}(x^p) = \overline{f}^p$ for $\overline{f} \in \mathbb{F}_p[x]$.

III.1.1 Proposition.
(1) $\Phi_n \in \mathbb{Z}[x]$.
(2) *For a prime number p one has*

$$\Phi_p = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

(3) *For any integer $n \geq 1$ one has*

$$\Phi_n = \prod_{1 \leq j < n,\, \gcd(j,n)=1} (x - \zeta_n^j).$$

This means that the zeros of Φ_n are precisely all elements of order n in $\mathbb{C}^\times$. Moreover, the degree of Φ_n is $\phi(n)$, where ϕ is Euler's phi-function defined by $\phi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times$.

Proof. (1) follows from the Lemma of Gauss applied to $f = X^n - 1$ and $g = \Phi_n$.
(2) follows from (3), however, we now give a direct proof. Write $f(x) = x^{p-1} + \ldots + x + 1$ and make the substitution $x = 1 + t$. Then

$$f(1+t) = \frac{(1+t)^p - 1}{(1+t) - 1} = \sum_{i=1}^{p} \binom{p}{i} t^{i-1}$$

satisfies Eisenstein's criterion and is therefore irreducible. It follows that f itself is irreducible, and since $f = (x^p - 1)/(x-1)$, the number ζ_p is a zero of f. Hence $f = \Phi_p$ which is what we wanted to prove.
(3) $x^n - 1 = \prod_{1 \le j \le n}(x - \zeta^j)$ and Φ_n can only contain the factors $(x - \zeta^j)$ with $\gcd(j,n) = 1$. Indeed, $\gcd(j,n) = d > 1$ implies that ζ^j is a root of $x^{n/d} - 1$. From $\Phi_n(\zeta^j) = 0$ it would follow that Φ_n divides $x^{n/d} - 1$ which leads to the contradiction $\zeta^{n/d} = 1$.

In order to see that every $x - \zeta^j$ with $\gcd(j,n) = 1$ is a factor of Φ_n, we use a trick. Decompose j as a product $p_1 \cdots p_t$ of (not necessarily distinct) prime factors. The p_i do not divide n since $\gcd(j,n) = 1$. Write $\zeta^j = (\cdots((\zeta^{p_1})^{p_2})\cdots)^{p_t}$. We claim that the following statement holds:

$$(*)\ \textit{if } p \textit{ does not divide } n \textit{ and if } \Phi(\alpha) = 0, \textit{ then } \Phi(\alpha^p) = 0.$$

Using this assertion one finds

$$\Phi_n(\zeta) = 0 \Rightarrow \Phi_n(\zeta^{p_1}) = 0 \Rightarrow \Phi_n(\zeta^{p_1 p_2}) = 0 \Rightarrow \cdots \Rightarrow \Phi_n(\zeta^j) = 0.$$

We will prove (*) by deriving a contradiction from the assumptions: $p \nmid n$ and $\Phi_n(\alpha) = 0$ and $\Phi_n(\alpha^p) \neq 0$.

The equality $x^n - 1 = \Phi_n \cdot f$ with $f(\alpha^p) = 0$ is clear. Now $f(\alpha^p) = 0$ means that α is a zero of $f(x^p)$. Therefore Φ_n divides $f(x^p)$. According to the Lemma of Gauss, this division takes place in the ring $\mathbb{Z}[x]$. Hence $\Phi_n(x^p)\Phi_n(x)$ divides $x^{pn} - 1$ in the ring $\mathbb{Z}[x]$. After reduction modulo p, one finds that $\overline{\Phi}_n^{p+1}$ divides $x^{pn} - 1 = (x^n - 1)^p$ in $\mathbb{F}_p[x]$. However $p \nmid n$ and $x^n - 1$ has only simple roots. The multiplicity of any root of $(x^n - 1)^p$ is p and this is the contradiction that we wanted to find. ∎

The proof of the following corollary is left to the reader.

III.1.2 Corollary. *The Galois group* $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ *is isomorphic to the group* $(\mathbb{Z}/n\mathbb{Z})^\times$. *This isomorphism identifies the unit* $a \bmod n$ *with the field automorphism sending* ζ_n *to* ζ_n^a.

III.1.3 Proposition (Formulas for Φ_n)**.**
1. $x^n - 1 = \prod_{d|n} \Phi_d$ *(i.e., the product over all divisors* d *of* n*).*
2. $\Phi_n = \prod_{d|n}(x^d - 1)^{\mu(n/d)}$*, where* μ *is the Möbius function given by* $\mu(1) = 1$, $\mu(n) = 0$ *if* n *contains a square* $\neq 1$ *and* $\mu(p_1 \cdots p_t) = (-1)^t$ *if* $p_1, \ldots, p_t$ *are distinct primes.*
3. $\Phi_{np}(x) = \Phi_n(x^p)$ *if the prime* p *divides* n.
4. $\Phi_{np}(x) = \Phi_n(x^p)\Phi_n(x)^{-1}$ *if the prime* p *does not divide* n.

Proof. (1) follows since both polynomials are monic and by Proposition III.1.1(3) have the same zeros.

(2) is a special case of the "Möbius inversion". This is the statement

$$\textit{If for all } n \ge 1 \textit{ the formula } f_n = \prod_{d|n} g_d \textit{ holds, then}$$

$$g_n = \prod_{d|n} f_d^{\mu(n/d)} \textit{holds for all } n \ge 1.$$

In the proof of this inversion formula one uses the easily deduced formulas $\sum_{d|n} \mu(d) = 0$ for $n > 1$ and $\sum_{d|n} \mu(d) = 1$ for $n = 1$.
The proofs of 3. and 4. are left as an exercise. ∎

III.2 An application to tangent values

Take $n \in \mathbb{Z}_{\geq 3}$ and consider $t_n := \tan(\pi/n)$. With $\zeta := \zeta_{4n} = e^{2\pi i/4n}$ we have $i = \zeta^n$ and

$$t_n = \frac{\sin(\pi/n)}{\cos(\pi/n)} = \frac{\zeta^2 - \zeta^{-2}}{\zeta^n(\zeta^2 + \zeta^{-2})} \in \mathbb{Q}(\zeta).$$

We will compute the degree $[\mathbb{Q}(t_n) : \mathbb{Q}]$ by determining the subgroup of the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong (\mathbb{Z}/4n\mathbb{Z})^\times$ corresponding to the intermediate field $\mathbb{Q}(t_n) \subset \mathbb{Q}(\zeta)$.

To do so, first note that $\overline{a} \in (\mathbb{Z}/4n\mathbb{Z})^\times$ yields the automorphism σ_a of $\mathbb{Q}(\zeta)$ defined by $\zeta \mapsto \zeta^a$. Its effect on t_n is given as

$$t_n \mapsto \sigma_a(t_n) = \frac{\zeta^{2a} - \zeta^{-2a}}{\zeta^{an}(\zeta^{2a} + \zeta^{-2a})} = \begin{cases} \tan(\pi a/n) & \text{if} \quad a \equiv 1 \bmod 4; \\ -\tan(\pi a/n) & \text{if} \quad a \equiv 3 \bmod 4. \end{cases}$$

So σ_a is the identity when restricted to $\mathbb{Q}(t_n)$ precisely when

$$\text{either } \overline{a} \in (\mathbb{Z}/4n\mathbb{Z})^\times \text{ satisfies } a \equiv 1 \bmod 4 \text{ and } \tan(\pi/n) = \tan(\pi a/n),$$

$$\text{or } \overline{a} \in (\mathbb{Z}/4n\mathbb{Z})^\times \text{ satisfies } a \equiv 3 \bmod 4 \text{ and } \tan(\pi/n) = \tan(-\pi a/n).$$

Solving these tangent equalities results in

$$\begin{cases} a \equiv 1 \bmod n \\ a \equiv 1 \bmod 4 \end{cases} \quad \text{or} \quad \begin{cases} a \equiv -1 \bmod n \\ a \equiv -1 \bmod 4 \end{cases}$$

The number of solutions $\overline{a} \in (\mathbb{Z}/4n\mathbb{Z})^\times$ now depends on the parity of n:
(i). If $n \equiv 0 \bmod 4$ then one finds the condition $a \equiv \pm 1 \bmod n$. This gives precisely 8 pairwise distinct values $\overline{a}$, namely

$$\overline{a} \in \left\{\overline{1}, \overline{-1}, \overline{n+1}, \overline{n-1}, \overline{2n+1}, \overline{2n-1}, \overline{3n+1}, \overline{3n-1}\right\}.$$

They form a subgroup of order 8 in $(\mathbb{Z}/4n\mathbb{Z})^\times \cong \mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ which in fact by the Galois correspondence equals $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q}(t_n))$. So in this case by Corollary I.3.4 we have $[\mathbb{Q}(\zeta) : \mathbb{Q}(t_n)] = 8$ and therefore using Proposition III.1.1(c) we have

$$[\mathbb{Q}(t_n) : \mathbb{Q}] = [\mathbb{Q}(\zeta) : \mathbb{Q}]/8 = \phi(4n)/8 = \phi(n/2).$$

(ii). If $n \equiv 2 \bmod 4$ then we obtain the condition $a \equiv \pm 1 \bmod 2n$, and this yields exactly 4 distinct values $\overline{a}$, namely

$$\overline{a} \in \left\{\overline{1}, \overline{-1}, \overline{2n+1}, \overline{2n-1}\right\}.$$

Reasoning as in the previous case (note that now $n/2$ is odd) one concludes

$$[\mathbb{Q}(t_n) : \mathbb{Q}] = \phi(4n)/4 = \phi(n) = \phi(n/2).$$

(iii). Finally, if $2 \nmid n$ then only $a \equiv \pm 1 \bmod 4n$ solve the desired equations. So one obtains the subgroup $\{\overline{1}, \overline{-1}\} \subset (\mathbb{Z}/4n\mathbb{Z})^\times$. As a consequence one concludes for n odd

$$[\mathbb{Q}(t_n) : \mathbb{Q}] = \phi(4n)/2 = \phi(n).$$

Now consider the special case where $n = p$ is an odd prime number. The calculation presented above shows that the minimal polynomial of $t_p = \tan(\pi/p)$ over $\mathbb{Q}$ has degree $[\mathbb{Q}(t_p) : \mathbb{Q}] = \phi(p) = p - 1$. To construct this minimal polynomial, put $\alpha := \pi/p$. One has $(e^{\alpha i})^p = -1$, hence $(\cos(\alpha) + i\sin(\alpha))^p + 1 = 0$. Dividing by $(\cos(\alpha))^p$ and taking the imaginary part of the resulting expression yields

$$\sum_{j=0}^{(p-1)/2} \binom{p}{2j+1} (-1)^j t_p^{2j+1} = 0.$$

So t_p is a zero of

$$\sum_{j=0}^{(p-1)/2} \binom{p}{2j+1} (-1)^j x^{2j} \in \mathbb{Z}[x].$$

As this has degree $p-1$ and leading coefficient $(-1)^{(p-1)/2}$, it is up to at most a sign the minimal polynomial of t_p over $\mathbb{Q}$.

Example: t_3 is a zero of x^2-3 and t_5 is a zero of x^4-10x^2+5 and t_7 is a zero of $x^6-21x^4+35x^2-7$.

III.3 Quadratic reciprocity

In this section $p, q \in \mathbb{Z}_{>0}$ are prime numbers.

Quadratic reciprocity relates the problem whether X^2-q has a zero in $\mathbb{F}_p$ to the 'reciprocal' problem whether X^2-p has a zero in $\mathbb{F}_q$. In the proof presented here, no Galois theory will be used, but we do use roots of unity and certain so-called Gauss sums. We will briefly mention a property of such Gauss sums in terms of Galois theory.

III.3.1 Definition. Let p be an odd prime and let $a \in \mathbb{Z}$.

The *Legendre-symbol* $\left(\frac{a}{p}\right) \in \{-1, 0, 1\}$ is defined by:

$$\left(\tfrac{a}{p}\right) = \quad 0 \quad \Leftrightarrow \quad a \text{ is divisible by } p,$$

$$\left(\tfrac{a}{p}\right) = \quad 1 \quad \Leftrightarrow \quad \exists x \in \mathbb{Z} : x \not\equiv 0 \bmod p \text{ and } x^2 \equiv a \bmod p,$$

$$\left(\tfrac{a}{p}\right) = -1 \quad \Leftrightarrow \quad \nexists x \in \mathbb{Z} : x^2 \equiv a \bmod p.$$

So in particular: the polynomial X^2-a has a zero in $\mathbb{F}_p \iff \left(\frac{a}{p}\right) \in \{+1, 0\}$.

III.3.2 Example. In $\mathbb{F}_{11}$ it holds that:

$$(\pm 1)^2 = 1, \quad (\pm 2)^2 = 4, \quad (\pm 3)^3 = 9, \quad (\pm 4)^2 = 5, \quad (\pm 5)^2 = 3,$$

so by the definition of the Legendre-symbol one finds:

$$\left(\frac{a}{11}\right) = 1 \quad \text{if } a \equiv 1, 3, 4, 5, 9 \bmod 11.$$

The remaining $10/2 = 5$ elements of $\mathbb{F}_{11}^\times$ are not squares:

$$\left(\frac{a}{11}\right) = -1 \quad \text{for } a \equiv 2, 6, 7, 8, 10 \bmod 11.$$

III.3.3 Theorem. *Let $a \in \mathbb{Z}$. Identifying $-1, 0, 1 \in \mathbb{Z}$ with $-\overline{1}, \overline{0}, \overline{1} \in \mathbb{F}_p$ it holds that*

$$\left(\frac{a}{p}\right) = \overline{a}^{\frac{p-1}{2}} \qquad (\in \mathbb{F}_p).$$

Moreover for all $n, m \in \mathbb{Z}$ one has

$$\left(\frac{nm}{p}\right) = \left(\frac{n}{p}\right)\left(\frac{m}{p}\right).$$

Proof. If $\overline{a} = \overline{0} \in \mathbb{F}_p$ then $\left(\frac{a}{p}\right) = 0$ and $\overline{a}^{\frac{p-1}{2}} = \overline{0}$, showing the first claim in this case.

If $\overline{a} \neq \overline{0}$ then $\overline{a}^{p-1} = 1$ since $\mathbb{F}_p^\times$ is a group consisting of $p-1$ elements. Hence $\overline{a}^{\frac{p-1}{2}}$ is a zero of $X^2 - 1 = 0$ and therefore $\overline{a}^{\frac{p-1}{2}} \in \{\overline{1}, \overline{-1}\}$, as these are the only zeros of $X^2 - 1$ in the field $\mathbb{F}_p$. We can therefore define

$$\epsilon : \mathbb{F}_p^\times \longrightarrow \{\overline{1}, \overline{-1}\} \qquad x \mapsto x^{\frac{p-1}{2}}.$$

Note that ϵ is a homomorphism of (multiplicative) groups. Since $\mathbb{F}_p^\times$ is a cyclic group is, $y \in \mathbb{F}_p^\times$ exists with $y^{\frac{p-1}{2}} \neq \overline{1}$. So ϵ is surjective and $\#\mathrm{Ker}(\epsilon) = \frac{p-1}{2}$.

In case $\overline{a} = x^2$ for some $x \in \mathbb{F}_p$, $x \neq 0$, then $\overline{a}^{\frac{p-1}{2}} = x^{p-1} = 1$, so $\overline{a} \in \mathrm{Ker}(\epsilon)$. As $\overline{a}^2 = \overline{b}^2 \Leftrightarrow \overline{a} = \pm\overline{b}$, one obtains in this way $\frac{p-1}{2}$ elements of $\mathrm{Ker}(\epsilon)$. Now $\#\mathrm{Ker}(\epsilon) = \frac{p-1}{2}$ and therefore

$$\mathrm{Ker}(\epsilon) = \{\overline{a} \in \mathbb{F}_p^\times : \overline{a} = x^2 \text{ for some } x \in \mathbb{F}_p^\times\}.$$

As a consequence $\epsilon(\overline{b}) = \overline{-1}$ in case b is not a square modulo p, so:

$$\left(\frac{a}{p}\right) = \epsilon(\overline{a})$$

for all $a \in \mathbb{Z}$ such that $\overline{a} \neq \overline{0} \in \mathbb{F}_p$. This proves the first assertion.

The second assertion is evident in case $\overline{n} = \overline{0}$ or $\overline{m} = \overline{0}$. In all other cases one uses that the Legendre-symbol agrees with the homomorphism ϵ, as was observed before. This finishes the proof. ∎

III.3.4 Example. In $\mathbb{F}_{11}$ it holds that

$$2^5 = 32 = -1, \qquad 3^5 = 27 \cdot 9 = 5 \cdot 9 = 1, \qquad 5^5 = 125 \cdot 25 = 4 \cdot 3 = 1,$$

hence 2 is not a square modulo 11, while both 3 and 5 are squares modulo 11 (see also the previous Example III.3.2). ▬

III.3.5 Remark. Given an odd prime p, it is not difficult to compute $(-1)^{\frac{p-1}{2}}$. Namely, write p as $p = 4k+1$ or as $p = 4k+3$ for an integer k. Then

$$(-1)^{\frac{p-1}{2}} = 1 \quad \text{if} \quad p = 4k+1, \qquad (-1)^{\frac{p-1}{2}} = -1 \quad \text{if} \quad p = 4k+3.$$

Hence the polynomial $X^2 + 1$ has a zero in $\mathbb{F}_p$, (here p is an odd prime) if and only if $p \equiv 1 \bmod 4$. (Note that for $p = 2$ one has $X^2 + 1 = (X+1)^2$ in $\mathbb{F}_2$.)

Of course the same conclusions can be obtained by using the well-known fact that $\mathbb{F}_p^\times$ is a cyclic group consisting of $p-1$ elements, so an element of order 4 (i.e., an element whose square is $-1 \neq 1$) exists, if and only if $4|(p-1)$.

III.3.6 Definition. Given a field K and a positive integer n, an element $\zeta \in K$ is called a *primitive n-th root of unity* if ζ has order n in the multiplicative group $K^\times$. (In other words, $\zeta^n = 1$ and $\zeta^m \neq 1$ for $1 \leq m < n$).

Note that using Proposition III.1.1(1), the n-th cyclotomic polynomial Φ_n introduced in Section III.1 can be regarded as a polynomial in $K[x]$ for any field K. From Proposition III.1.3(1) it then follows that any primitive n-th root of unity in K is in fact a zero of the polynomial $\Phi_n \in K[x]$.

However, a zero in K of Φ_n is not necessarily a primitive n-th root of unity. For example if $\mathrm{char}(K) = p > 0$ then $x^p - 1 = (x-1)^p$ in $K[x]$, hence primitive p-th roots of unity do not exist in K.

III.3.7 Example. If $\mathrm{char}(K) \neq 2$ then -1 is the only primitive 2-nd root of unity in K.

For $K = \mathbb{C}$ the primitive p-th roots of unity (p a prime number) are precisely the $p-1$ (pairwise distinct) complex numbers

$$\zeta_p^k \quad (1 \leq k \leq p-1), \qquad \text{with} \quad \zeta_p := \cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p}.$$

In particular $\frac{-1\pm i\sqrt{3}}{2}$ are the two primitive 3-rd roots of unity. ∎

III.3.8 Example. The primitive 3-rd roots of unity in $\mathbb{F}_7$ are 2 and 4, as follows from $2^3 = 8 = 1$ and $4^3 = 64 = 1$ in $\mathbb{F}_7$.

Note that $\mathbb{F}_{25} \cong \mathbb{F}_5[X]/(X^2-2)$, since X^2-2 has no zeros in $\mathbb{F}_5$. Therefore every element of $\mathbb{F}_{25}$ can be expressed uniquely as $a + b\alpha$ with $a, b \in \mathbb{F}_5$ and $\alpha^2 = 2$. The primitive 3-rd roots of unity in $\mathbb{F}_{25}$ are therefore $3(-1 \pm \alpha)$, namely

$$(3(-1+\alpha))^3 = 2(-1+3\alpha-3\alpha^2+\alpha^3) = 2(-1+3\alpha-1+2\alpha) = 1,$$

and similarly for $3(-1-\alpha)$. ∎

We will now study primitive p-th roots of unity, for p an odd prime number, somewhat more extensively. In particular we introduce Gauss-sums and use them to obtain information concerning quadratic equations (over finite fields).

III.3.9 Definition. Let p be an odd prime number and let K be a field of characteristic $\mathrm{char}(K) \neq p$. Suppose $\zeta \in K$ is a primitive p-th root of unity.

The map $\mathbb{Z} \to K^\times$ given by $n \mapsto \zeta^n$ is a homomorphism of groups with kernel $p\mathbb{Z} \subset \mathbb{Z}$. Therefore one obtains a well-defined homomorphism $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p \to K^\times$ given by

$$x \mapsto \zeta^x := \zeta^n, \qquad \text{for} \quad x = \overline{n} = n + p\mathbb{Z} \in \mathbb{F}_p.$$

Similarly the Legendre symbol $\left(\frac{n}{p}\right) \in \{0, +1, -1\}$ can, as before, be regarded as an element of K. Since it depends only on p and on n mod p, we obtain a map $\mathbb{F}_p \to K$ given by

$$x \mapsto \left(\frac{x}{p}\right) := \left(\frac{n}{p}\right) \in K, \qquad \text{for} \quad x = \overline{n} = n + p\mathbb{Z} \in \mathbb{F}_p.$$

The *Gauss-sum* $\tau \in K$ is defined as

$$\tau := \sum_{x \in \mathbb{F}_p} \left(\frac{x}{p}\right) \cdot \zeta^x.$$

III.3.10 Example. Take $p = 3$. Considering $K = \mathbb{C}$ and $\zeta = \frac{-1+i\sqrt{3}}{2}$, one finds

$$\tau = \left(\frac{0}{3}\right)\cdot\zeta^0 + \left(\frac{1}{3}\right)\cdot\zeta^1 + \left(\frac{2}{3}\right)\cdot\zeta^2 = 0 + \zeta - \zeta^2 = i\sqrt{3}.$$

Note that with the choice $\zeta = \frac{-1-i\sqrt{3}}{2}$ one obtains $\tau = -i\sqrt{3}$, the complex conjugate of the Gauss-sum above. So the Gauss-sum depends on the choice of the primitive p-th root of unity ζ, however for our application this is not important.

For $K = \mathbb{F}_7$ and 3-rd root of unity $\zeta = 2 \in \mathbb{F}_7$ we have

$$\tau = \left(\frac{0}{3}\right)\cdot 2^0 + \left(\frac{1}{3}\right)\cdot 2^1 + \left(\frac{2}{3}\right)\cdot 2^2 = 2 - 4 = 5.$$

Note that $\tau^2 = 5^2 = -3$ in $\mathbb{F}_7$, analogous to the case $K = \mathbb{C}$. ∎

III.3.11 Theorem. *Suppose p is an odd prime number and K is a field with $\mathrm{char}(K) \neq p$ and $\zeta \in K$ is a primitive p-th root of unity. The associated Gauss-sum τ satisfies*

$$\tau^2 = \left(\frac{-1}{p}\right) \cdot p.$$

Proof. One computes

$$\begin{aligned}
\tau^2 &= \left(\sum_{x\in\mathbb{F}_p} \left(\frac{x}{p}\right)\cdot\zeta^x\right)\cdot\left(\sum_{y\in\mathbb{F}_p}\left(\frac{y}{p}\right)\cdot\zeta^y\right)\\
&= \sum_{x,y\in\mathbb{F}_p}\left(\frac{xy}{p}\right)\cdot\zeta^{x+y}\\
&= \sum_{z\in\mathbb{F}_p}\left(\sum_{x\in\mathbb{F}_p}\left(\frac{x(z-x)}{p}\right)\right)\cdot\zeta^z \qquad (\text{met } z = x+y).
\end{aligned}$$

For $x = 0$ one finds $\left(\frac{x(z-x)}{p}\right) = 0$, and for $x \neq 0$ the following holds:

$$\begin{aligned}
\left(\tfrac{x(z-x)}{p}\right) &= \left(\tfrac{-x^2}{p}\right)\cdot\left(\tfrac{1-zx^{-1}}{p}\right)\\
&= \left(\tfrac{-1}{p}\right)\left(\tfrac{x^2}{p}\right)\cdot\left(\tfrac{1-zx^{-1}}{p}\right)\\
&= \left(\tfrac{-1}{p}\right)\cdot\left(\tfrac{1-zx^{-1}}{p}\right).
\end{aligned}$$

Hence

$$\tau^2 = \left(\frac{-1}{p}\right)\cdot\sum_{z\in\mathbb{F}_p} c_z\zeta^z \qquad \text{with} \quad c_z = \sum_{x\in\mathbb{F}_p^\times}\left(\frac{1-zx^{-1}}{p}\right).$$

The case $z = 0$ leads to

$$c_0 = \sum_{x\in\mathbb{F}_p^\times}\left(\frac{1}{p}\right) = \sum_{x\in\mathbb{F}_p^\times} 1 = p-1.$$

In case $z \neq 0$, and x running over the elements of $\mathbb{F}_p^\times$, one has that zx^{-1} runs over $\mathbb{F}_p^\times$ as well, and $w := 1 - zx^{-1}$ runs over $\mathbb{F}_p - \{1\}$. This implies

$$c_z = \left(\sum_{w\in\mathbb{F}_p}\left(\frac{w}{p}\right)\right) - \left(\frac{1}{p}\right) = 0 - 1 = -1 \qquad (\text{with} \quad z \in \mathbb{F}_p^\times),$$

where it is used that $\mathbb{F}_p$ contains as many elements w such that $\left(\frac{w}{p}\right) = 1$ as it contains elements with $\left(\frac{w}{p}\right) = -1$. Substituting the values of c_z in the formula for τ^2 now yields

$$\begin{aligned}
\tau^2 &= \left(\frac{-1}{p}\right)\cdot\left(p - 1 + \sum_{z\in\mathbb{F}_p^\times}(-1)\zeta^z\right)\\
&= \left(\frac{-1}{p}\right)\cdot\left(p - 1 - \sum_{z\in\mathbb{F}_p^\times}\zeta^z\right)\\
&= \left(\frac{-1}{p}\right)\cdot\left(p - \sum_{z=0}^{p-1}\zeta^z\right)\\
&= \left(\frac{-1}{p}\right)\cdot p,
\end{aligned}$$

where we also used that

$$\sum_{z=0}^{p-1}\zeta^z = \frac{\zeta^p - 1}{\zeta - 1} = 0.$$

This proves Theorem III.3.11. ∎

III.3.12 Remark. We "explain" in terms of Galois theory why in some sense the formula defining a Gauss-sum is very natural. So let p be an odd prime, and $\zeta = \zeta_p = e^{2\pi i/p} \in \mathbb{C}$. The field $K := \mathbb{Q}(\zeta)$ is a Galois extension of $\mathbb{Q}$ with Galois group $G \cong (\mathbb{Z}/p\mathbb{Z})^\times$ (see Corollary III.1.2). The element $a \bmod n \in (\mathbb{Z}/p\mathbb{Z})^\times$ corresponds to the automorphism of K given by $\zeta \mapsto \zeta^a$.

The subgroup $H \subset G$ consisting of all squares has index 2 in G. Hence the field $M := K^H$ which by the Galois correspondence is associated to the subgroup H, is a quadratic extension of $\mathbb{Q}$. We describe a generator of M. This means: we want an element $\alpha \in K$ with the properties $\sigma(\alpha) = \alpha$ for all $\sigma \in H$ (this guarantees that $\alpha \in M$) and moreover $\tau(\alpha) \neq \alpha$ for some $\tau \in G \setminus H$ (this guarantees $\alpha \notin \mathbb{Q}$).

An obvious first choice is

$$\alpha := \sum_{x \in \mathbb{F}_p^\times \text{ a square}} \zeta^x.$$

Namely, if $\sigma \in H$, then $\sigma(\zeta) = \zeta^y$ for some square $y \in \mathbb{F}_p^\times$, hence

$$\sigma(\alpha) = \sum_{x \in \mathbb{F}_p^\times \text{ a square}} \zeta^{xy} = \alpha$$

(since when x runs over the squares in $\mathbb{F}_p^\times$, so does xy). Moreover when $\nu \in G \setminus H$ then $\nu(\zeta) = \zeta^z$ for some $z \in \mathbb{F}_p^\times$ which is *not* a square. Then

$$\nu(\alpha) = \sum_{x \in \mathbb{F}_p^\times \text{ a square}} \zeta^{xz} = \sum_{w \in \mathbb{F}_p^\times \text{ not a square}} \zeta^w.$$

Now observe that

$$\sum_{x \in \mathbb{F}_p^\times \text{ a square}} \zeta^x + \sum_{w \in \mathbb{F}_p^\times \text{ not a square}} \zeta^w = \sum_{j=1}^{p-1} \zeta^j = -1.$$

As a consequence $\nu(\alpha) = -1 - \alpha$. To show that $\alpha \notin \mathbb{Q}$ we therefore need to establish that $\alpha \neq -1 - \alpha$, in other words, that $2\alpha \neq -1$.

A short argument for this (different and slightly more elementary reasonings are given later as well as in the exercises) is the following. Evaluating at ζ yields a surjective ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}[\zeta]$ with kernel $\mathbb{Z}[x] \cdot \Phi_p$. Hence we have $\mathbb{Z}[\zeta] \cong \mathbb{Z}[x]/(\Phi_p)$. Now let $\mathbb{F}_q$ be the splitting field of $X^p - 1$ over $\mathbb{F}_2$, and take a primitive p-th root of unity $\overline{\zeta} \in \mathbb{F}_q$. Then via $\mathbb{Z}[x] \to \mathbb{F}_q$, $f(x) \mapsto f(\overline{\zeta})$ one obtains a ring homomorphism $\mathbb{Z}[\zeta] \to \mathbb{F}_q$. It sends $\alpha \in \mathbb{Z}[\zeta]$ to some element $\overline{\alpha} \in \mathbb{F}_q$ and then 2α to $2\overline{\alpha} = 0$, since $\mathrm{char}(\mathbb{F}_q) = 2$. So $2\alpha \neq -1$ because the image of -1 is not $0 \in \mathbb{F}_q$.

We now have that $M = K^H = \mathbb{Q}[\alpha]$, a quadratic extension of $\mathbb{Q}$ with basis (as $\mathbb{Q}$-vectorspace) $1, \alpha$ and with Galois group over $\mathbb{Q}$ generated by the restriction to M of any $\nu \in G$ as above. On the basis $1, \alpha$ such ν is given by the matrix $\left(\begin{smallmatrix} 1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$. So a basis for M over $\mathbb{Q}$ consisting of eigenvectors of ν is $1, 2\alpha + 1$. We have

$$2\alpha + 1 = 2\left(\sum_{x \in \mathbb{F}_p^\times \text{ a square}} \zeta^x\right) - \left(\sum_{j=1}^{p-1} \zeta^j\right) = \tau,$$

precisely our Gauss-sum! It is an eigenvector of ν with eigenvalue -1, and hence $\nu(\tau^2) = (\nu(\tau))^2 = (-\tau)^2 = \tau^2$, showing that $\tau^2 \in \mathbb{Q}$. Theorem III.3.11 shows a little more, namely it determines τ^2 as an element of $\mathbb{Q}$, namely $(2\alpha + 1)^2 = \tau^2 = \pm p$, providing another argument why $2\alpha \neq -1$.

Theorem III.3.11 describes a zero $\tau \in K$, in case $\mathrm{char}(K) \neq p$ and K contains a primitive p-th root of unity, of the polynomial $X^2 - \left(\frac{-1}{p}\right)p$. For example, we can take as K the field $\Omega_{\mathbb{F}_q}^{X^p-1}$, for $p \neq q$ two distinct primes (and $p \neq 2$). In the latter case a natural question is, whether τ is contained in the field $\mathbb{F}_q$. In other words, whether or not $\tau^q = \tau$.

III.3.13 Lemma. *Suppose p and q are distinct odd primes. Let K be a field with $char(K) = q$, containing a primitive p-th root of unity. Then*

$$\tau^q = \left(\frac{q}{p}\right)\cdot\tau, \qquad \textit{and in particular} \qquad \tau \in \mathbb{F}_q \iff \left(\frac{q}{p}\right) = 1.$$

Proof. By the definition of Gauss-sum and the fact that $x \mapsto x^q$ is a field homomorphism of K, it follows that

$$\begin{aligned}
\tau^q &= \left(\sum_{x\in\mathbb{F}_p}\left(\frac{x}{p}\right)\cdot\zeta^x\right)^q \\
&= \sum_{x\in\mathbb{F}_p}\left(\frac{x}{p}\right)\cdot\zeta^{qx} \\
&= \sum_{y\in\mathbb{F}_p}\left(\frac{yq^{-1}}{p}\right)\cdot\zeta^{y} \qquad\qquad (\text{with} \quad y = x\overline{q} \in \mathbb{F}_p) \\
&= \left(\frac{q^{-1}}{p}\right)\cdot\sum_{y\in\mathbb{F}_p}\left(\frac{y}{p}\right)\cdot\zeta^{y} \\
&= \left(\frac{q}{p}\right)\cdot\tau,
\end{aligned}$$

here we used $\overline{q} = \overline{q}^2\cdot\overline{q}^{-1}$, hence $\overline{q}^{-1}$ is a square in $\mathbb{F}_p^\times$ precisely when $\overline{q}$ is.

For the last assertion, note that $\tau \neq 0$ by Theorem III.3.11. This proves the lemma. ∎

III.3.14 Theorem. (quadratic reciprocity) (Gauss, 1801).
Suppose p and q are distinct odd primes. Then

$$\begin{aligned}
\left(\frac{q}{p}\right) &= \left(\frac{p}{q}\right) \qquad \text{if } p \equiv 1 \bmod 4 \text{ or } q \equiv 1 \bmod 4, \\
\left(\frac{q}{p}\right) &= -\left(\frac{p}{q}\right) \qquad \text{if } p \equiv q \equiv 3 \bmod 4.
\end{aligned}$$

Formulated somewhat shorter:

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Moreover

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}, \qquad \textit{and} \qquad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

Proof. First, note that $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$ is immediate from Theorem III.3.3.

We now show quadratic reciprocity. Take $K = \Omega_{\mathbb{F}_q}^{X^p-1}$ and let τ be the Gauss-sum as before. By Lemma III.3.13 we have

$$\tau^{q-1} = \left(\frac{q}{p}\right).$$

On the other hand from Theorem III.3.11 we obtain

$$\tau^{q-1} = (\tau^2)^{\frac{q-1}{2}} = \left(\frac{-1}{p}\right)^{\frac{q-1}{2}} p^{\frac{q-1}{2}}.$$

Now Theorem III.3.3 implies

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}, \qquad p^{\frac{q-1}{2}} = \left(\frac{p}{q}\right) \qquad (\in \mathbb{F}_q).$$

Combining the above equalities one concludes

$$\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}\left(\frac{p}{q}\right),$$

which is evidently equivalent to quadratic reciprocity.

Finally we will show for any odd prime p that

$$(\exists x \in \mathbb{F}_p : \quad x^2 = 2) \iff p \equiv \pm 1 \bmod 8.$$

Given an odd prime p, the polynomial X^4+1 has a zero ζ in some finite extension K of $\mathbb{F}_p$. So $\zeta^4 = -1$ and $\zeta^8 = 1$. Put $x := \zeta + \zeta^7 \in K$, then

$$x^2 = (\zeta+\zeta^7)^2 = \zeta^2 + \zeta^6 + 2 = 2,$$

where we used $\zeta^6 = \zeta^4\zeta^2 = -\zeta^2$.

In case $p = 8k+1$ or $p = 8k+7$ then

$$\left.\begin{array}{llll} \zeta^p = \zeta, & \zeta^{7p} & = & \zeta^7 \quad \text{if } p \equiv 1 \bmod 8 \\ \zeta^p = \zeta^7, & \zeta^{7p} & = & \zeta \quad \text{if } p \equiv 7 \bmod 8 \end{array}\right\} \Longrightarrow x^p = (\zeta+\zeta^7)^p = \zeta+\zeta^7 = x,$$

where we used that $x \mapsto x^p$ is a field homomorphism. From $x^p = x$ it follows that $x \in \mathbb{F}_p$. This shows that in case $p \equiv \pm 1 \bmod 8$ an $x \in \mathbb{F}_p$ exists with $x^2 = 2$, which means $\left(\frac{2}{p}\right) = 1$.

In case $p = 8k+3$ or $p = 8k+5$ then (using $\zeta^4 = -1$) one finds

$$\left.\begin{array}{lll} \zeta^p & = & \zeta^3 = -\zeta^7 \quad \text{if } p \equiv 3 \bmod 8 \\ \zeta^p & = & \zeta^5 = -\zeta \quad \text{if } p \equiv 5 \bmod 8 \end{array}\right\} \Longrightarrow x^p = (\zeta+\zeta^7)^p = -(\zeta+\zeta^7) = -x.$$

Hence the zero x of X^2-2 is not in $\mathbb{F}_p$. The other zero (which is $-x$) is then not in $\mathbb{F}_p$ either. Hence $\left(\frac{2}{p}\right) = -1$.

This completes the proof of Theorem III.3.14. ∎

III.3.15 Example. We present a few simple applications of Theorem III.3.14.

$$\left(\frac{7}{11}\right) = -\left(\frac{11}{7}\right) = -\left(\frac{4}{7}\right) = -1,$$

since $4 = 2^2$. So 7 is not a square modulo 11. Next,

$$\left(\frac{5}{11}\right) = \left(\frac{11}{5}\right) = \left(\frac{1}{5}\right) = 1,$$

so 5 is a square modulo 11. Indeed, $4^2 = 16 = 5$ in $\mathbb{F}_{11}$. Using that the Legendre-symbol is multiplicative (see III.3.3), we find:

$$\left(\frac{6}{11}\right) = \left(\frac{2}{11}\right)\left(\frac{3}{11}\right) = (-1)\cdot -\left(\frac{11}{3}\right) = (-1)\cdot -\left(\frac{2}{3}\right) = -1.$$

∎

III.4 Exercises

1. Give a proof of Corollary III.1.2.

2. Prove parts (3) and (4) of Proposition III.1.3.

3. Let $\lambda \in \mathbb{Q}$. Prove that $\cos(2\pi\lambda)$ is algebraic over $\mathbb{Q}$. Prove that $\mathbb{Q}(\cos(2\pi\lambda))$ is a Galois extension. Determine its Galois group.

4. Determine Φ_{72} in a handy way.

5. Show for any odd prime p that the minimal polynomial of $\tan(\pi/p)$ over $\mathbb{Q}$ is an Eisenstein polynomial.

6. Find all integers $n \geq 3$ such that $\tan(\pi/n) \in \mathbb{Q}$.

7. For each of the integers $n \geq 3$ such that $\tan(\pi/n)$ has degree 2 over $\mathbb{Q}$, find the corresponding minimal polynomial.

8. Consider $f := x^6 - 21x^4 + 35x^2 - 7 \in \mathbb{Z}[x]$. As we saw in Section III.2, this is the minimal polynomial of $t_7 = \tan(\pi/7)$ over $\mathbb{Q}$, and $\mathbb{Q}(t_7)/\mathbb{Q}$ is a Galois extension.
   (a) Describe all zeros of f as values of the tangent function.
   (b) Writing f as a product of factors of degree 1 and comparing coefficients, what identities between tangent values do you obtain?
   (c) Show using Galois theory that $\mathbb{Q}(t_7)$ contains a unique quadratic field $\mathbb{Q}(\sqrt{d})$. (in fact, it holds that $\sqrt{7} \in \mathbb{Q}(t_7)$. Factor f over $\mathbb{Q}(\sqrt{7})$ and deduce more tangent indentities from this!)

9. Show for primes $p > 3$ that: 3 is a square modulo $p \quad \Leftrightarrow \quad p \equiv 1, -1 \bmod 12$.

10. Show for $p > 3$ prime that -3 is a square in $\mathbb{F}_p$ if and only if $p \equiv 1 \bmod 3$.

11. Let $p > 2$ be prime and $n \in \mathbb{Z}$. Prove that if $p|(n^2 - 2)$, then $p \equiv \pm 1 \bmod 8$.

12. Compute the Legendre-symbols $\left(\frac{5}{101}\right)$, $\left(\frac{6}{101}\right)$, $\left(\frac{7}{101}\right)$, $\left(\frac{11}{101}\right)$.

13. In Remark III.3.12, two arguments are provided showing that (with p an odd prime and ζ a primitive p-th root of unity in $\mathbb{C}$) one has $\alpha := \sum \overline{a} \in \mathbb{F}_p^{\times 2} \zeta^a \neq -1/2$. In this exercise we give two more proofs of this fact.
   (a) Use that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p-1$ and show that every subset of $\{1, \zeta, \zeta^2, \zeta^3, \ldots, \zeta^{p-1}\}$ consisting of $p-1$ elements, is a basis of $\mathbb{Q}(\zeta)$ as a vectorspace over $\mathbb{Q}$. Conclude from this that $\alpha \notin \mathbb{Q}$.
   (b) Another proof: forgetting the multiplication in the ring $\mathbb{Z}[\zeta] \cong \mathbb{Z}[X]/(\Phi_p)$, one obtains the finitely generated abelian group $(\mathbb{Z}[\zeta], +, 0)$. Use this to show that $(\mathbb{Z}[\alpha], +, 0)$ is finitely generated as well, and conclude from this that $\alpha \notin \mathbb{Q} \setminus \mathbb{Z}$.

14. Let p be a prime number and let $\zeta \in \mathbb{C}^\times$ be a primitive p-th root of unity. Put $K = \mathbb{Q}(\zeta)$. Show that a unique field M exists with $\mathbb{Q} \subset M \subset K$ and $[M : \mathbb{Q}] = 2$. Present an example of a cyclotomic field $\mathbb{Q}(\zeta_n)$ containing more than one field M with $[M : \mathbb{Q}] = 2$.

IV SYMMETRIC POLYNOMIALS

IV.1 Definition and results

Let R be a commutative ring with 1, and n an integer ≥ 1.

IV.1.1 Definition. A polynomial $f \in R[X_1, X_2, ..., X_n]$ is called *symmetric* if f is fixed under *all* permutations of $X_1, X_2, ..., X_n$.

IV.1.2 Examples. The polynomials

$$\sum_{i=1}^{n} X_i, \qquad \prod_{i=1}^{n} X_i, \qquad \sum_{i=1}^{n} X_i^k \qquad (\text{with } k \in \mathbb{Z}_{\geq 0})$$

are symmetric in $R[X_1, \ldots, X_n]$.

The polynomial $X_1X_2 + X_2X_3 + X_3X_4 + X_4X_1$ is *not* symmetric: it is not fixed under interchanging X_1 and X_2 (here $n = 4$).

If Z is a new variable, then the polynomial

$$(Z - X_1)(Z - X_2)\cdots(Z - X_n) \quad \in R[X_1, X_2, ..., X_n][Z]$$

can be written as

$$Z^n - \sigma_1 Z^{n-1} + \sigma_2 Z^{n-2} - ... + (-1)^{n-1}\sigma_{n-1}Z + (-1)^n \sigma_n$$

with

$$\begin{aligned}
\sigma_1 &= X_1 + X_2 + ... + X_n \\
\sigma_2 &= X_1X_2 + X_1X_3 + ... + X_1X_n + X_2X_3 + ... + X_{n-1}X_n, \\
\sigma_3 &= X_1X_2X_3 + ... = \textstyle\sum_{1 \leq i < j < k \leq n} X_iX_jX_k, \\
&\vdots \\
\sigma_t &= \textstyle\sum_{1 \leq i_1 < i_2 < ... < i_t \leq n} X_{i_1}X_{i_2}\cdots X_{i_t}, \\
&\vdots \\
\sigma_n &= X_1X_2\cdots X_n.
\end{aligned}$$

The coefficients $\sigma_1, \sigma_2, ..., \sigma_n$ here are symmetric polynomials called the *elementary symmetric polynomials*. From $\sigma_1, \sigma_2, ..., \sigma_n$ one obtains other symmetric polynomials by addition, multiplication, and multiplication by elements of R.

IV.1.3 Example. Take $n = 2$, then

$$\sigma_1 = X_1 + X_2, \quad \sigma_2 = X_1 X_2.$$

Other symmetric polynomials are

$$\sigma_1^2 = X_1^2 + 2X_1X_2 + X_2^2, \quad \sigma_1^2 - 2\sigma_2 = X_1^2 + X_2^2, \quad \sigma_1^3 - 3\sigma_1\sigma_2 = X_1^3 + X_2^3, \quad \text{etcetera.}$$

—■

One observes that every polynomial in $\sigma_1, \sigma_2, ..., \sigma_n$ with coefficients from R, so every $g(\sigma_1, \sigma_2, \ldots, \sigma_n)$ for $g \in R[X_1, \ldots, X_n]$, is a symmetric polynomial. The converse of this holds as well:

IV.1.4 Theorem. (Main theorem of symmetric polynomials) *Any symmetric polynomial $f \in R[X_1, X_2, ..., X_n]$ can be written as a polynomial in $\sigma_1, \sigma_2, ..., \sigma_n$ with coefficients in R (so $f(X_1, \ldots, X_n) = g(\sigma_1, \ldots, \sigma_n)$ for some $g \in R[X_1, \ldots, X_n]$).*

Moreover this way of writing f is unique (so g is uniquely determined by f).

Proof. Let $f \neq 0$ be symmetric. Order the nonzero terms $rX_1^{a_1}X_2^{a_2}\cdots X_n^{a_n}$ appearing in f in such a way that a term $r \cdot X_1^{a_1}X_2^{a_2}\cdots X_n^{a_n}$ appears 'before' $r' \cdot X_1^{b_1}X_2^{b_2}\cdots X_n^{b_n}$ if $a_i > b_i$ for the least i with $a_i \neq b_i$ ('lexicographical ordering').

The 'leading term'

$$rX_1^{c_1}X_2^{c_2}\cdots X_n^{c_n} \qquad (r \in R,\ r \neq 0)$$

of f then has

$$\begin{aligned} c_1 &= (\text{largest } a_1 \text{ appearing in } f \text{ as exponent of } X_1), \\ c_2 &= (\text{largest } a_2 \text{ apprearing with a given } a_1 = c_1), \end{aligned}$$

et cetera. We call r the *leading coefficient* of f.

As f is symmetric, $c_1 \geq c_2 \geq ... \geq c_n$ since otherwise interchanging two X_i's can provide an 'earlier' term of f.

We claim that the symmetrical polynomial

$$r\sigma_1^{c_1-c_2}\sigma_2^{c_2-c_3}\cdots\sigma_{n-1}^{c_{n-1}-c_n}\sigma_n^{c_n}$$

also has leading term $rX_1^{c_1}X_2^{c_2}\cdots X_n^{c_n}$. Indeed,

$$\begin{aligned} &\sigma_1 \text{ has leading term} && X_1, \\ &\sigma_2 \text{ has leading term} && X_1X_2 \\ &\vdots \\ &\sigma_n \text{ has leading term} && X_1X_2\cdots X_n. \end{aligned}$$

Now the rule

$$\text{leading term}\,(g) \cdot \text{leading term}\,(h) = \text{leading term}\,(g \cdot h)$$

(which holds for polynomials g, h with leading coefficient 1), shows

$$\begin{aligned} \text{leading term}\,(\sigma_1^{c_1-c_2}\sigma_2^{c_2-c_3}\ldots\sigma_n^{c_n}) &= X_1^{c_1-c_2}\cdot(X_1X_2)^{c_2-c_3}\cdots(X_1X_2\cdots X_n)^{c_n} \\ &= X_1^{c_1}X_2^{c_2}\cdots X_n^{c_n}, \end{aligned}$$

as claimed.

So one concludes that

$$f_1 := f - r\sigma_1^{c_1-c_2}\cdots\sigma_n^{c_n}$$

only contains terms which lexicographically come later than the leading term of f.

If $f_1 = 0$ then we found an expression for f as desired:

$$f = r\sigma_1^{c_1-c_2}\cdots\sigma_n^{c_n}.$$

If $f_1 \neq 0$, note that f_1 is symmetric as well, so we can apply the same reasoning to f_1 as was done for f. This gives

$$f_2 = f_1 - r'\sigma_1^{c_1'-c_2'}\cdots\sigma_n^{c_n'}$$

and all terms appearing in f_2 come lexicographically later than the leading term $r'X_1^{c_1'}\cdots X_n^{c_n'}$ of f_1. In case $f_2 = 0$ we are done:

$$f = r\sigma_1^{c_1-c_2}\cdots\sigma_n^{c_n} + r'\sigma_1^{c_1'-c_2'}\cdots\sigma_n^{c_n'},$$

otherwise we continue with f_2.

It remains to show that this process terminates, so in the sequence $f_1, f_2, f_3, \ldots$ one has $f_k = 0$ for some k.

To this end one introduces the *total degree* $\mathrm{tdeg}(f)$ of f, which is the maximal $a_1 + a_2 + \cdots + a_n$ one obtains from the terms $r \cdot X_1^{a_1}\cdots X_n^{a_n} (\neq 0)$ of f. Note that $\mathrm{tdeg}(\sigma_i) = i$, and therefore

$$\begin{aligned}\mathrm{tdeg}(\sigma_1^{c_1-c_2}\cdots\sigma_n^{c_n}) &= 1\cdot(c_1-c_2) + 2\cdot(c_2-c_3) + \ldots + n\cdot c_n \\ &= c_1 + c_2 + \ldots + c_n \\ &\leq \mathrm{tdeg}(f).\end{aligned}$$

It follows that

$$\mathrm{tdeg}(f_1) \leq \mathrm{tdeg}(f)$$

and more generally

$$\ldots \leq \mathrm{tdeg}(f_m) \leq \mathrm{tdeg}(f_{m-1}) \leq \ldots \leq \mathrm{tdeg}(f)$$

However, for a given total degree only finitely many products $X_1^{a_1}\cdots X_n^{a_n}$ are possible. In every step of the process at least one such product is erased and all remaining ones come lexicographically later. This shows that after finitely many steps we have $f_k = 0$, completing the proof of the first assertion in Theorem IV.1.4.

It remains to show: if $g_1 \neq g_2$ are polynomals in n variables over R, then $g_1(\sigma_1, \sigma_2, \ldots, \sigma_n) \neq g_2(\sigma_1, \sigma_2, \ldots, \sigma_n)$. Writing $g = g_1 - g_2$ one concludes that it suffices to show:

$$\text{if} \quad g \in R[Y_1, \ldots, Y_n],\ g \neq 0, \quad \text{then} \quad g(\sigma_1, \sigma_2, \ldots, \sigma_n) \neq 0.$$

Every term appearing in g can be written as

$$rY_1^{a_1-a_2}\cdot Y_2^{a_2-a_3}\cdots Y_n^{a_n},$$

with $r \in R$, $r \neq 0$, $a_i \in \mathbb{Z}_{\geq 0}$. Consider the term in g such that the corresponding $a_1, a_2, \ldots, a_n$ (i.e., the product $X_1^{a_1}X_2^{a_2}\cdots X_n^{a_n}$) comes first in the lexicographical ordering. Substituting σ_i for Y_i one obtains a polynomial in $X_1, \ldots, X_n$ with leading term

$$(*) \qquad rX_1^{a_1}X_2^{a_2}\cdots X_n^{a_n},$$

since the other terms $r'\sigma^{a_1'-a_2'}\ldots\sigma_n^{a_n'}$ give rise to polynomials in $X_1, \ldots, X_n$ with a later leading term. Hence $(*)$ does not cancel against other terms, and therefore $g(\sigma_1, \sigma_2, \ldots, \sigma_n) \neq 0$. This proves the theorem. ∎

IV.1.5 Remark. In the case $R = K$ is a field, we discuss some relation between Theorem IV.1.4 and Galois theory. Namely, put $L := K(X_1,\ldots,X_n)$, the field of rational functions over K in the variables $X_1,\ldots X_n$. Then the elementary symmetric polynomials σ_j are elements of L, so we obtain a subfield $M := K(\sigma_1,\ldots,\sigma_n)$ of L. Any permutation of the variables $X_1,\ldots,X_n$ extends to a K-linear automorphism of L and these permutations in fact restrict to the identity map on the subfield M. Moreover L is the splitting field over M of the separable polynomial $\prod_{j=1}^{n}(X-X_j) \in M[X]$. As this polynomial has degree n, we have $[L:M] \leq n!$.

The Galois group $\mathrm{Gal}(L/M)$ contains the subgroup corresponding to all permutations of $X_1,\ldots,X_n$, so in particular $[L:M] = \#\mathrm{Gal}(L/M) \geq n!$.

The conclusion is that $[L:M] = n!$ and $G := \mathrm{Gal}(L/M) \cong S_n$. In particular every permutation of the variables X_j extends to an element of G, and all elements of G are obtained in this way. Galois theory then tells us that $L^G = M$, so the rational functions in the X_j's that are symmetric, which means invariant under any permutation of the variables, are precisely the rational functions in the σ_j's.

Theorem IV.1.4, which we proved without any reference to Galois theory, states that this even holds when we replace 'rational functions' by 'polynomials', and even stronger, even when we replace the field K by an arbitrary ring R.

IV.1.6 Example. Take $n = 3$, and

$$f = X_1^3X_2 + X_1^3X_3 + X_1X_2^3 + X_1X_3^3 + X_2^3X_3 + X_2X_3^3.$$

The terms here are already lexicographically ordered, and the leading term $X_1^3X_2$ has $c_1 = 3,$, $c_2 = 1$, and $c_3 = 0$. Following the proof of Theorem IV.1.4 we must subtract from f the symmetric expression

$$\sigma_1^{c_1-c_2}\sigma_2^{c_2-c_3}\sigma_3^{c_3} = \sigma_1^2\sigma_2 = (X_1+X_2+X_3)^2\cdot(X_1X_2+X_1X_3+X_2X_3)$$

$$= X_1^3X_2 + X_1^3X_3 + 2X_1^2X_2^2 + 5X_1^2X_2X_3 + 2X_1^2X_3^2 + X_1X_2^3$$
$$+5X_1X_2^2X_3 + 5X_1X_2X_3^2 + X_1X_3^3 + X_2^3X_3 + 2X_2^2X_3^2 + X_2X_3^3.$$

The result is

$$f_1 = -2X_1^2X_2^2 - 5X_1^2X_2X_3 - 2X_1^2X_3^2 - 5X_1X_2^2X_3 - 5X_1X_2X_3^2 - 2X_2^2X_3^2.$$

From this one subtracts

$$-2\sigma_2^2 = -2X_1^2X_2^2 - 4X_1^2X_2X_3 - 2X_1^2X_3^2 - 4X_1X_2^2X_3 - 4X_1X_2X_3^2 - 2X_2^2X_3^2,$$

which leads to

$$f_2 = f_1 - (-2\sigma_2^2) = -X_1^2X_2X_3 - X_1X_2^2X_3 - X_1X_2X_3^2.$$

Subtracting from this $-\sigma_1\sigma_3$ one obtains 0, so

$$f = \sigma_1^2\sigma_2 - 2\sigma_2^2 - \sigma_1\sigma_3.$$

Hence the g as mentioned in Theorem IV.1.4 is $X_1^2X_2 - 2X_2^2 - X_1X_3$. ∎

Usually Theorem IV.1.4 is applied to the following situation. Let $f \in R[X_1,\ldots,X_n]$ be symmetric, and take $\alpha_1, \alpha_2, \ldots, \alpha_n \in R$. By Theorem IV.1.4 f can be expressed in $\sigma_1,\sigma_2,\ldots,\sigma_n$, hence $f(\alpha_1,\alpha_2,\ldots,\alpha_n)$ can be expressed in

$$\begin{aligned}
&\sigma_1(\alpha_1,\ldots,\alpha_n) = \alpha_1 + \ldots + \alpha_n,\\
&\sigma_2(\alpha_1,\ldots,\alpha_n) = \alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_{n-1}\alpha_n,\\
&\vdots\\
&\sigma_n(\alpha_1,\ldots,\alpha_n) = \alpha_1\alpha_2\ldots\alpha_n,
\end{aligned}$$

which are precisely ± the coefficients of

$$(X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n).$$

Roughly speaking this means that every symmetric expression in 'the n zeros' of a monic polynomial of degree n in one variable, can be expressed in the coefficients of this polynomial.

This assertion becomes particularly relevant of these n zeros are not contained in the ring R itself, but only in some extension $R' \supset R$. We now first present an example and then continue with a general result.

IV.1.7 Example. Consider $h = X^3 - X - 1 \in \mathbb{Z}[X]$. In $\mathbb{Z}$, and even in $\mathbb{Q}$, no zero of h exists (as follows using the lemma of Gauss). However, $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$ exist such that

$$X^3 - X - 1 = (X-\alpha_1)(X-\alpha_2)(X-\alpha_3).$$

Comparing coefficients yields

$$\begin{array}{lll} \sigma_1(\alpha_1,\alpha_2,\alpha_3) & = \alpha_1+\alpha_2+\alpha_3 & = \quad 0 \\ \sigma_2(\alpha_1,\alpha_2,\alpha_3) & = \alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3 & = -1 \\ \sigma_3(\alpha_1,\alpha_2,\alpha_3) & = \alpha_1\alpha_2\alpha_3 & = \quad 1. \end{array}$$

Hence Theorem IV.1.4 implies:
is $f \in \mathbb{Z}[X_1,X_2,X_3]$ an arbitrary symmetric polynomial, then $f(\alpha_1,\alpha_2,\alpha_3) \in \mathbb{Z}$ (although $\alpha_1,\alpha_2,\alpha_3 \notin \mathbb{Z}$).

Taking as f the polynomial from Example IV.1.6, so

$$f = X_1^3X_2 + X_1^3X_3 + \ldots + X_2X_3^3 = \sigma_1^2\sigma_2 - 2\sigma_2^2 - \sigma_1\sigma_3$$

one finds by substiting $X_i := \alpha_i$:

$$\alpha_1^3\alpha_2 + \alpha_1^3\alpha_3 + \ldots + \alpha_2\alpha_3^3 = 0^2\cdot(-1) - 2\cdot(-1)^2 - 0.1 = -2.$$

■

IV.1.8 Theorem. *Let R' be a commutative ring (with 1) and $R \subset R$ a subring. Suppose $h \in R[X]$ has degree n and suppose $\alpha_1,\alpha_2,\ldots,\alpha_n \in R'$ exist such that*

$$h = (X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n).$$

Then for every symmetric $f \in R[X_1,X_2,\ldots,X_n]$ it holds that

$$f(\alpha_1,\alpha_2,\ldots,\alpha_n) \in R.$$

Proof. For all i we have

$$\sigma_i(\alpha_1,\ldots,\alpha_n) \in R$$

since this is up to sign the coefficient of X^i in $h \in R[X]$. As the symmetric f can be written as $f = g(\sigma_1,\ldots,\sigma_n)$ for some $g \in R[X_1,\ldots,X_n]$ (see Theorem IV.1.4), substituting the α_i in $g(\sigma_1,\ldots,\sigma_n)$ yields an element in R, as desired. ■

IV.1.9 Example. An important symmetric polynomial is

$$D = \prod_{1\le i<j\le n} (X_i - X_j)^2.$$

The *discriminant* of a polynomial

$$h = X^n + a_1X^{n-1} + \ldots + a_n = (X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n)$$

is defined as

$$\Delta(h) = \prod_{1\le i<j\le n} (\alpha_i - \alpha_j)^2 = D(\alpha_1, \alpha_2, ..., \alpha_n).$$

This discriminant can be expressed in $a_1, a_2, ..., a_n$. For $n = 2, 3, 4$ one obtains the following formulas:

$$\Delta(X^2 + aX + b) = a^2 - 4b,$$

$$\Delta(X^3 + aX^2 + bX + c) = a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc,$$

$$\Delta(X^4 + aX^3 + bX^2 + cX + d) =$$

$$\frac{1}{27}\{4(b^2 - 3ac + 12d)^3 - (2b^3 - 72bd + 27a^2d - 9abc + 27c^2)^2\}$$

(when expanding the latter expression, it turns out that the 27 in the denominator cancels).

Given a (unitary) ring R, the discriminant of $h = X^n + a_1X^{n-1} + ... + a_n \in R[X]$ is defined as an expression in the a_j's, regardless whether or not h can be factored in $R[X]$ as a product of n factors $(X - \alpha_i)$.

The use of the discriminant relies on the observarion that in the case where the ring R is an integral domain, we have

$$\Delta(h) = 0 \iff \exists i, j \quad i \neq j : \quad \alpha_i = \alpha_j,$$

in other words, the discriminant is zero if and only if the polynomial has a multiple zero. The case $n = 2$ is already discussed in high school. An other application of the discriminant is discussed in Exercise 5 on page 34.

Is K a field of characteristic $\mathrm{char}(K) \neq 3$ (so that $\frac{1}{3} \in K$), then the substitution $X := X - \frac{1}{3}a$ brings a polynomial $f = X^3 + aX^2 + bX + c = \prod(X - \alpha_i)$ of degree 3 into the form $g = X^3 + pX + q$. Note that $\triangle(f) = \triangle(g)$, since the zeros of g are $\beta_i := \alpha_i + \frac{1}{3}a$ and $\alpha_i - \alpha_j = \beta_i - \beta_j$. The discriminant of g is simply $\triangle(g) = -(4p^3 + 27q^2)$. ∎

IV.1.10 Example. We will use symmetric polynomials to obtain expressions for the zeros of a polynomial of degree 3.

Let K be a field of characteristic $\mathrm{char}(K) \neq 2, 3$, and let $f \in K[X]$ be a monic polynomial of degree 3:

$$f = X^3 + aX^2 + bX + c.$$

Let $\alpha_1, \alpha_2, \alpha_3$ be the zeros of f (in a splitting field of f over K, see Section I.3). Then

$$\begin{aligned} -a &= \alpha_1 + \alpha_2 + \alpha_3, \\ b &= \alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3, \\ -c &= \alpha_1\alpha_2\alpha_3. \end{aligned}$$

Let ω be a primitive third root of unity (in an extension of the chosen splitting field of f over K), so $\omega \neq 1$, $\omega^3 = 1$). Put

$$\begin{aligned} A_1 &:= \alpha_1 + \omega\alpha_2 + \omega^2\alpha_3, \\ A_2 &:= \alpha_1 + \omega^2\alpha_2 + \omega\alpha_3 \end{aligned}$$

We consider what happens to A_i when permuting the α_i's. For $\rho := (123) \in S_3$ we find

$$\begin{aligned} \rho = (123): \quad A_1 &\mapsto \alpha_2 + \omega\alpha_3 + \omega^2\alpha_1 = \omega^2 A_1, \\ A_2 &\mapsto \alpha_2 + \omega^2\alpha_3 + \omega\alpha_1 = \omega A_2. \end{aligned}$$

Moreover,

$$\tau = (23): A_1 \mapsto A_2, \qquad (23): A_2 \mapsto A_1.$$

Since the group S_3 is generated by ρ and τ it follows that

$$A_1^3 + A_2^3, \qquad A_1A_2$$

are symmetric polynomials in $\alpha_1, \alpha_2, \alpha_3$.

Therefore by Theorem IV.1.4 they can be expressed in terms of the elementary symmetric polynomials σ_i which are in our case, up to sign, equal to the coefficients of f. A short calculation reveals

$$\begin{array}{rrl} 2B := & A_1^3 + A_2^3 & = -2a^3 + 9ab - 27c, \\ A := & A_1 A_2 & = a^2 - 3b. \end{array}$$

Since

$$(T - A_1^3)(T - A_2^3) = T^2 - 2BT + A^3,$$

we conclude that A_1^3, A_2^3 are given by

$$A_i^3 = B \pm \sqrt{B^2 - A^3}.$$

Hence

$$A_i = \sqrt[3]{B \pm \sqrt{B^2 - A^3}}$$

(here one has three choices for the cube root '$\sqrt[3]{\ }$'). Finally one determines α_1 by noting that

$$\begin{array}{rcl} 3\alpha_1 & = & (\alpha_1 + \alpha_2 + \alpha_3) + (\alpha_1 + \omega\alpha_2 + \omega^2\alpha_3) + (\alpha_1 + \omega^2\alpha_2 + \omega\alpha_3) \\ & = & -a + A_1 + A_2, \end{array}$$

where it is used that $\omega^2 + \omega + 1 = 0$.

As an explicit example, take

$$f = X^3 + 2X^2 - X - 2 \quad \in \mathbb{Q}[X].$$

Here

$$B = 10, \quad A = 7,$$

$$A_1 = \sqrt[3]{10 + \sqrt{10^2 - 7^3}} = \sqrt[3]{10 + 9i\sqrt{3}}$$

(we chose a + sign, this turns out to be irrelevant). There are three solutions in $\mathbb{C}$ to $A_1^3 = 10 + 9i\sqrt{3}$, namely

$$A_1 = -2 + i\sqrt{3} \quad \text{and} \quad A_1 = \frac{1}{2}(-1 + 3i\sqrt{3}) \quad \text{and} \quad A_1 = \frac{1}{2}(5 + i\sqrt{3}).$$

Since $A_1 A_2 = A = 7$ the corresponding A_2's are

$$A_2 = -2 - i\sqrt{3}, \quad A_2 = \frac{1}{2}(-1 + 3i\sqrt{3}), \quad A_2 = \frac{1}{2}(5 - i\sqrt{3}).$$

We now obtain the three zeros α of f using $\alpha = \frac{1}{3}(-a + A_1 + A_2)$, they are

$$-2, \quad -1, \quad 1.$$

The formulas for the roots of cubic equations which are obtained in this way, were in a different way found by Cardano en Tartaglia around 1540. They are called the *Cardano formulas*. ∎

IV.1.11 Remark. We briefly return to the described solution method for solving cubic equations. We interpret the method in terms of Galois theory. Namely, let K be a field (of characteristic $\neq 3, \neq 2$) and let $\sigma_1, \sigma_2, \sigma_3$ be variables. To be in exactly the situation discussed before, we will assume that K contains a primitive third root of unity ω. Over $K(\sigma_1, \sigma_2, \sigma_3)$ we consider the polynomial $f = X^3 - \sigma_1 X^2 + \sigma_2 X - \sigma_3$.

Let L be a splitting field of f over $K(\sigma_1,\sigma_2,\sigma_3)$ and let X_1,X_2,X_3 be the three zeros of f in L. As we saw in Remark IV.1.5 $L = K(X_1,X_2,X_3)$ and the extension $K(X_1,X_2,X_3) \supset K(\sigma_1,\sigma_2,\sigma_3)$ is Galois, with Galois group S_3 acting by permutations on X_1,X_2,X_3.

The elements $A_1^3,A_2^3 \in K(X_1,X_2,X_3)$ used in deriving the Cardano formulas are invariant under the subgroup $A_3 \subset S_3$, while the 2-cycles in S_3 interchange A_1^3 and A_2^3. So $A_1^3,A_2^3 \notin K(\sigma_1,\sigma_2,\sigma_3)$ but they are in the field L^{A_3} of invariants under A_3. This is a quadratic extension of $K(\sigma_1,\sigma_2,\sigma_3)$, and indeed we found a quadratic polynomial over $K(\sigma_1,\sigma_2,\sigma_3)$ with the A_j^3's as zeros: $T^2-2BT+A^3$. The element $A_1 \in L$ is not in this subfield since it is not fixed under the automorphisms in A_3. However A_1 satisfies the cubic equation $A_i^3 = W$ for some given $W \in K(\sigma_1,\sigma_2,\sigma_3)(A_1^3)$.

So we found the zeros of f by constructing a quadratic extension of $K(\sigma_1,\sigma_2,\sigma_3)$ and then taking the cube root of a suitable element in that extension.

In Exercise 4 a similar idea is used to obtain the zeros of a polynomial f of degree 4, which we now briefly sketch. In this case, the 'general' Galois group is S_4. We first describe 3 combinations C_i of the zeros of f which are invariant under the normal subgroup

$$H := \{e, (12)(34), (13)(24), (14)(23)\} \subset S_4.$$

Since $S_4/H \cong S_3$ the C_i generate a Galois extension with Galois group S_3. The polynomial $(X-C_1)(X-C_2)(X-C_3)$ then has coefficients which are invariant under all of S_4, so they can be expressed in the coefficients of f. Since it turns out to be relatively easy to obtain the zeros of f in terms of square roots of the C_i, this reduces the problem of finding expressions for the zeros of f to the problem of finding such expressions for a cubic polynomial, and that was done before.

Unfortunately, for $n > 4$ the group A_n is the only nontrivial normal subgroup of S_n. For this reason, solving an equation of degree $n > 4$ is in an essential way more difficult than the cases $n \leq 4$.

IV.2 Exercises

1. Express the symmetric polynomial $X_1^3+X_2^3+X_3^3$ (here $n=3$) in terms of $\sigma_1,\sigma_2,\sigma_3$.

2. In the proof of IV.1.4 we used the rule

$$\text{leadingterm}(g)\cdot\text{leadingterm}(h)=\text{leadingterm}(g\cdot h)$$

for polynomials g,h with leading coefficient 1. Show that this rule is false in general if we allow g,h to have *zero divisors* as leading coefficents.

3. Let $(X-\alpha_1)(X-\alpha_2)(X-\alpha_3)=X^3-X-1$, with $\alpha_1,\alpha_2,\alpha_3\in\mathbb{C}$. Put $s_k=\alpha_1^k+\alpha_2^k+\alpha_3^k$ for $k\in\mathbb{Z}$. Show:

$$\begin{array}{ll} s_{-1}=-1,\ s_0=3, & s_1=0\\ s_k=s_{k-2}+s_{k-3} & \text{for all } k\in\mathbb{Z},\\ s_k\in\mathbb{Z} & \text{for all } k\in\mathbb{Z}\ \text{(also negative ones!)}. \end{array}$$

4. Take $f=X^4+aX^3+bX^2+cX+d\ \in K[X]$ where K is a field of characteristic $\text{char}(K)\neq 2,3$. Write $\alpha_1,\ldots,\alpha_4$ for the zeros of f in some extension of K, so $f=\prod(X-\alpha_j)$.

   (a) Define

$$\begin{aligned} C_1 &= (\alpha_1+\alpha_2-\alpha_3-\alpha_4)^2\\ C_2 &= (\alpha_1-\alpha_2+\alpha_3-\alpha_4)^2\\ C_3 &= (\alpha_1-\alpha_2-\alpha_3+\alpha_4)^2. \end{aligned}$$

   Express α_1 in terms of $\sqrt{C_i}$ and the coefficient a of f.

   (b) Verify that the S_4 action (permuting the α_i's) also permutes the C_i's. Check that the subgroup H in Remark IV.1.11 leaves all C_i invariant.

   (c) Show that

$$\begin{aligned} C_1+C_2+C_3 &= 3a^2-8b\\ C_1C_2+C_1C_3+C_2C_3 &= 3a^4-16a^2b+16b^2+16ac-64d\\ C_1C_2C_3 &= (a^3-4ab+8c)^2. \end{aligned}$$

   (d) Explain how the information above can be used to find an expression for the solutions of a general equation of degree 4.

5. Suppose $f=X^3+aX^2+bX+c\in\mathbb{Q}[X]$ is an irreducible polynomial with zeros $\alpha_1,\alpha_2,\alpha_3\in\mathbb{C}$. The splitting field $\Omega_{\mathbb{Q}}^f\cong\mathbb{Q}(\alpha_1,\alpha_2,\alpha_3)$, hence

$$\sqrt{\Delta}:=(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)(\alpha_2-\alpha_3)\in\Omega_{\mathbb{Q}}^f,\quad\text{and}\quad\Delta\in\mathbb{Q},$$

   where Δ is the discriminant of f introduced in Example IV.1.9.

   (a) Prove that $[\Omega_{\mathbb{Q}}^f:\mathbb{Q}]=3$ or 6.

   (b) Prove that $\sqrt{\Delta}\notin\mathbb{Q}\Rightarrow[\Omega_{\mathbb{Q}}^f:\mathbb{Q}]=6$.

   (c) Now assume $\sqrt{\Delta}\in\mathbb{Q}$. Put $f=(X-\alpha_1)(X^2+rX+s)\in\mathbb{Q}(\alpha_1)[X]$. Show that $\alpha_2\in\mathbb{Q}(\alpha_1)$ by expressing α_2 in terms of $\sqrt{\Delta},a,b,c,r,s,\alpha_1\in\mathbb{Q}(\alpha_1)$. Conclude that $[\Omega_{\mathbb{Q}}^f:\mathbb{Q}]=3\Leftrightarrow\sqrt{\Delta}\in\mathbb{Q}$.

   (d) Describe $\text{Gal}(\Omega_{\mathbb{Q}}^f/\mathbb{Q})$ (the answer depends on Δ being a square or not).

V ALGEBRAICALLY CLOSED FIELDS

V.1 Definition and examples

V.1.1 Definition. A field K is called *algebraically closed* if for every $f \in K[X]$ such that $f \notin K$, an $\alpha \in K$ exists such that $f(\alpha) = 0$.

One of the statements in the next result is that if K is algebraically closed, then every $f \in K[X]$, $f \neq 0$, splits in $K[X]$ as a product of factors of degree 1.

V.1.2 Theorem. *Let K be a field. The following statements are equivalent:*

(a) *K is algebraically closed;*
(b) *every irreducible polynomial in $K[X]$ has degree 1;*
(c) *the only algebraic extension L of K is $L = K$;*
(d) *for every monic $f \in K[X]$ elements $\alpha_1, \alpha_2, ..., \alpha_n \in K$ exist with $f = \prod_{i=1}^{n}(X - \alpha_i)$.*

Proof. (a) $\Rightarrow$ (b). If (a) holds, then a polynomial of degree > 1 has a zero, hence is not irreducible.
(b) $\Rightarrow$ (c). Suppose $L \supset K$ is algebraic. Then any $\alpha \in L$ has a minimal polynomial f_K^α which is irreducible in $K[X]$, so $\deg(f_K^\alpha) = 1$ using (b). Hence $\alpha \in K$, which implies $L = K$.
(c) $\Rightarrow$ (d). The splitting field Ω_K^f of f over K is algebraic over K, hence (c) implies $\Omega_K^f = K$, which is what we wanted to show.
(d) $\Rightarrow$ (a). This is evident, since every α_i is a zero of f.
This completes the proof of Theorem V.1.2. ∎

V.1.3 Theorem. *Every algebraically closed field K is infinite.*

Proof. If K is finite, say $K = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, then

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

is not constant and has no zero in K. This proves V.1.3. ∎

The next result, originally proven in the doctoral dissertation (1799) of Carl Friedrich Gauss, is sometimes called the 'Fundamental Theorem of Algebra'.

V.1.4 Theorem. *The field $\mathbb{C}$ of complex numbers is algebraically closed.*

A simple analytic proof of this result, which relies on Liouville's theorem, is often discussed in introductory courses on complex function theory. Below we present a proof in which all ingredients except the following lemma are purely algebraic.

V.1.5 Lemma. *Let $f \in \mathbb{R}[X]$, and assume that $\beta, \gamma \in \mathbb{R}$ exist with $f(\beta) > 0$ and $f(\gamma) < 0$.*

Then $\alpha \in \mathbb{R}$ exists with $f(\alpha) = 0$.

Proof. (of Lemma V.1.5). This is a special case of the intermediate value theorem from Real Analysis, since polynomials define continous functions. ■

In the proof of Theorem V.1.4 we will make use of the following lemmas.

V.1.6 Lemma. *Any $f \in \mathbb{C}[X]$ of degree 2 has a zero in $\mathbb{C}$.*

Proof. We may assume that f is monic: $f = X^2 + \beta X + \gamma$, with $\beta, \gamma \in \mathbb{C}$. Writing

$$f = (X + \frac{1}{2}\beta)^2 - (\frac{1}{4}\beta^2 - \gamma)$$

it suffices to show that the complex number $\frac{1}{4}\beta^2 - \gamma$ has a square root in $\mathbb{C}$.

Write $\frac{1}{4}\beta^2 - \gamma = a + bi$, for $a, b \in \mathbb{R}$. First consider the case $b = 0$. If $a > 0$ one obtains the root $\sqrt{a}$ in $\mathbb{R}$ by applying Lemma V.1.5 to $g = X^2 - a$, observing that $g(0) < 0$, $g(a+1) > 0$. Is $a \leq 0$ then we have the square root $i\sqrt{|a|}$ in $\mathbb{C}$. This finishes the case $b = 0$.

From now on we assume $b \neq 0$ and we want $c, d \in \mathbb{R}$ such that

$$(c + di)^2 = (c^2 - d^2) + 2cdi = a + bi.$$

This means

$$c^2 - d^2 = a, \qquad 2cd = b.$$

As $b \neq 0$ the desired c, d should both be $\neq 0$ as well, and then

$$c = \frac{b}{2d}, \qquad \text{so} \quad \frac{b^2}{4d^2} - d^2 = a.$$

Therefore the real number d must be a zero of

$$g = 4X^4 + 4aX^2 - b^2 \in \mathbb{R}[X].$$

Since $g(0) < 0$ and $g(x) > 0$ for $x \in \mathbb{R}$ sufficiently large, Lemma V.1.5 implies that $d \in \mathbb{R}$ with $g(d) = 0$ exists. Then taking $c = \frac{b}{2d}$ shows that $a + bi$ has the square root $c + di$ in $\mathbb{C}$. This proves Lemma V.1.6. ■

V.1.7 Lemma. *Any $f \in \mathbb{R}[X]$ of odd degree has a zero in $\mathbb{R}$.*

Proof. We may assume that the leading coefficient of f is positive. Then $f(x) > 0$ for $x \in \mathbb{R}$ sufficiently large, and since $\deg(f)$ is odd we also have $f(x) < 0$ for $x \in \mathbb{R}$ sufficiently negative. Hence Lemma V.1.5 implies that f has a zero in $\mathbb{R}$. ■

V.1.8 Lemma. *Suppose that every nonconstant $f \in \mathbb{R}[X]$ (so with* real *coefficients) has a zero in $\mathbb{C}$. Then $\mathbb{C}$ is algebraically closed.*

Proof. Take $g = \sum_{i=0}^{n} a_i X^i \in \mathbb{C}[X]$, $g \notin \mathbb{C}$. We have to show that g has a zero in $\mathbb{C}$. Define

$$\bar{g} := \sum_{i=0}^{n} \bar{a}_i X^i \quad (\in \mathbb{C}[X])$$

where $\bar{a}_i$ denotes the complex conjugate of $a_i \in \mathbb{C}$. Put

$$f := g \cdot \bar{g} \quad (\in \mathbb{C}[X]).$$

Clearly all coefficients of f are real, so $f \in \mathbb{R}[X]$. Moreover $\deg(f) = 2 \cdot \deg(g)$ hence f is nonconstant. By the assumption of our lemma applied to f some $\alpha \in \mathbb{C}$ exists with $f(\alpha) = 0$. This means

$$g(\alpha) \cdot \bar{g}(\alpha) = 0.$$

Is $g(\alpha) = 0$ then we are done. In case $g(\alpha) \neq 0$ one concludes $\bar{g}(\alpha) = 0$, so

$$\sum_{i=0}^{n} \bar{a}_i \alpha^i = 0.$$

Taking the complex conjugate yields

$$\sum_{i=0}^{n} a_i \bar{\alpha}^i = 0,$$

so $g(\bar{\alpha}) = 0$ and again we found a zero in $\mathbb{C}$ of g. This proves Lemma V.1.8. ∎

We are now ready to prove the main result of this section.

Proof. (of Theorem V.1.4). By Lemma V.1.8 it suffices to show that any nonconstant $f \in \mathbb{R}[X]$ has a zero in $\mathbb{C}$. So take any such f. We may assume that f is monic. Put $n = \deg(f)$. We write $n = 2^k u$ with $k \in \mathbb{Z}_{\geq 0}$ and u an *odd* positive integer. The proof will be done by mathematical induction w.r.t. k.

If $k = 0$ then $\deg(f)$ is odd and Lemma V.1.7 shows that f has a zero in $\mathbb{C}$ (even in $\mathbb{R}$).

Now assume $k \geq 1$, so n is even. We use the field $L = \Omega_{\mathbb{C}}^{f} \supset \mathbb{C}$. In $L[X]$ we have

$$f = \prod_{i=1}^{n} (X - \alpha_i), \qquad \alpha_i \in L \quad (1 \leq i \leq n).$$

For any $c \in \mathbb{R}$ consider

$$g_c = \prod_{1 \leq i < j \leq n} (X - (\alpha_i + \alpha_j + c\alpha_i\alpha_j)) \qquad (\in L[X]).$$

The coefficients of g_c are symmetric polynomials in $\alpha_1, \alpha_2, \ldots, \alpha_n$, hence by Theorem IV.1.8 (applied to $R = \mathbb{R}$, $R' = L$) we have $g_c \in \mathbb{R}[X]$.

The degree of g_c equals the number of pairs (i, j) with $1 \leq i < j \leq n$, which is $\frac{1}{2}n(n-1) = 2^{k-1} \cdot u \cdot (n-1)$. Here $n-1$ is odd, so the number of factors 2 in $\deg(g_c)$ is $k-1$. Hence applying the induction hypothesis to g_c we conclude that g_c has a zero in $\mathbb{C}$. Now the zeros of g_c are the $\frac{1}{2}n(n-1)$ expressions $\alpha_i + \alpha_j + c\alpha_i\alpha_j$. We conclude: for every $c \in \mathbb{R}$ there exist i and j with $1 \leq i < j \leq n$ and $\alpha_i + \alpha_j + c\alpha_i\alpha_j \in \mathbb{C}$.

Here i and j depend on c. However there are only finitely many possibilities for i and j while we have infinitely many $c \in \mathbb{R}$. So a pair of real numbers $c \neq c'$ which give the same i, j. This means

$$\alpha_i + \alpha_j + c\alpha_i\alpha_j \in \mathbb{C}, \qquad \alpha_i + \alpha_j + c'\alpha_i\alpha_j \in \mathbb{C}.$$

Taking suitable linear combinations one finds that

$$\beta = \alpha_i + \alpha_j, \quad \gamma = \alpha_i\alpha_j \in \mathbb{C} \quad \text{andhence} \quad (X - \alpha_i)(X - \alpha_j) = X^2 - \beta X + \gamma \in \mathbb{C}[X].$$

Then Lemma V.1.6 tells us that this polynomial has a zero in $\mathbb{C}$, so $\alpha_i \in \mathbb{C}$ or $\alpha_j \in \mathbb{C}$. Hence f has a zero in $\mathbb{C}$, as desired.

This proves the induction step and therefore Theorem V.1.4. ∎

V.1.9 Corollary. *Every irreducible $f \in \mathbb{R}[X]$ has degree 1 or 2.*
A polynomial $X^2 + bX + c \in \mathbb{R}[X]$ is irreducible in $\mathbb{R}[X]$ if and only if $b^2 - 4c < 0$.

Proof. Let $f \in \mathbb{R}[X]$ be monic and irreducible, and let $\alpha \in \mathbb{C}$ be a zero of f. Then $f = f_{\mathbb{R}}^{\alpha}$, and

$$\deg(f) = [\mathbb{R}(\alpha) : \mathbb{R}] \leq [\mathbb{C} : \mathbb{R}] = 2.$$

This proves the first assertion in V.1.9. The second one follows from

$$X^2 + bX + c = (X + \frac{1}{2}b)^2 - \frac{1}{4}(b^2 - 4c)$$

and the fact that $X^2 - a$ is irreducible in $\mathbb{R}[X]$ if and only if $a < 0$. This shows V.1.9. ∎

V.2 The algebraic closure

V.2.1 Definition. An *algebraic closure* of a field K is a field extension $\overline{K} \supset K$ with the properties

i. $\overline{K}$ is algebraic over K;
ii. $\overline{K}$ is algebraically closed.

V.2.2 Example. From V.1.4 we know that $\mathbb{C}$ is an algebraic closure of of $\mathbb{R}$. ∎

V.2.3 Theorem. *The field $\mathbb{Q}$ of rational numbers has an algebraic closure.*

Proof. Put

$$\overline{\mathbb{Q}} = \{\alpha \in \mathbb{C} : \ \alpha \text{ is algebraic over } \mathbb{Q}\}.$$

From basic properties of the notion 'algebraic' we have that indeed $\overline{\mathbb{Q}}$ is a field. We verify condition (ii), which then implies that $\bar{\mathbb{Q}}$ is an algebraic closure of $\mathbb{Q}$.

Laat $f \in \overline{\mathbb{Q}}[X]$ be nonconstant. We need to show that f has a zero in $\overline{\mathbb{Q}}$. Using Theorem V.1.4 one obtains a zero α of f in $\mathbb{C}$. As $f \in \overline{Q}[X]$, this α is algebraic over $\overline{\mathbb{Q}}$. Moreover the field $\overline{\mathbb{Q}}$ is algebraic over $\mathbb{Q}$, and therefore $\overline{\mathbb{Q}}(\alpha)$ is an algebraic extension of $\mathbb{Q}$. In particular α is algebraic over $\mathbb{Q}$, so $\alpha \in \overline{\mathbb{Q}}$. This proves V.2.3. ∎

V.2.4 Remark. The algebraic closure of $\mathbb{Q}$ certainly does *not* equal $\mathbb{C}$, since (uncountably many) transcendental complex numbers exist.

More generally we have:

V.2.5 Theorem. *Every field K has an algebraic closure $\overline{K}$.*
Moreover $\overline{K}$ is unique up to K-isomorphism, meaning: if $\overline{K}_1$ and $\overline{K}_2$ are algebraic closures of K, then $\overline{K}_1 \cong_K \overline{K}_2$.

Proofs of this for the general case necessarily depend on "Zorn's Lemma", the same axiom in the foundations of set theory that was also needed to prove the existence of a maximal ideal in an arbitrary ring. For completeness, we copy here three proofs taken from J.S. Milne's online lecture notes *Fields and Galois Theory*, see www.jmilne.org/math/CourseNotes/FTc.pdf (Version 4.52, March 17, 2017, pages 86–88).

First proof of the existence of algebraic closures

(Bourbaki, Algèbre, Chap. V, §4.) An ***F-algebra*** is a ring containing F as a subring. Let $(A_i)_{i\in I}$ be a family of commutative F-algebras, and define $\bigotimes_F A_i$ to be the quotient of the F-vector space with basis $\prod_{i\in I} A_i$ by the subspace generated by elements of the form:

$(x_i)+(y_i)-(z_i)$ with $x_j+y_j=z_j$ for one $j\in I$ and $x_i=y_i=z_i$ for all $i\neq j$;

$(x_i)-a(y_i)$ with $x_j=ay_j$ for one $j\in I$ and $x_i=y_i$ for all $i\neq j$,

(ibid., Chap. II, 3.9). It can be made into a commutative F-algebra in an obvious fashion, and there are canonical homomorphisms $A_i\to\bigotimes_F A_i$ of F-algebras.

For each polynomial $f\in F[X]$, choose a splitting field E_f, and let $\Omega=(\bigotimes_F E_f)/M$ where M is a maximal ideal in $\bigotimes_F E_f$ (whose existence is ensured by Zorn's lemma). Note that $F\subset\bigotimes_F E_f$ and $M\cap F=0$. As Ω has no ideals other than (0) and Ω, it is a field (see 1.2). The composite of the F-homomorphisms $E_f\to\bigotimes_F E_f\to\Omega$, being a homomorphism of fields, is injective. Since f splits in E_f, it must also split in the larger field Ω. The algebraic closure of F in Ω is therefore an algebraic closure of F .

Second proof of the existence of algebraic closures

(Jacobson 1964, p144.) After (4.24) we may assume F to be infinite. This implies that the cardinality of every field algebraic over F is the same as that of F (ibid. p143). Choose an uncountable set Ξ of cardinality greater than that of F, and identify F with a subset of Ξ. Let S be the set of triples $(E,+,\cdot)$ with $E\subset\Xi$ and $(+,\cdot)$ a field structure on E such that $(E,+,\cdot)$ contains F as a subfield and is algebraic over it. Write $(E,+,\cdot)\le(E',+',\cdot')$ if the first is a subfield of the second. Apply Zorn's lemma to show that S has maximal elements, and then show that a maximal element is algebraically closed.

Third proof of the existence of algebraic closures

(Emil Artin.) Consider the polynomial ring $F[\dots,x_f,\dots]$ in a family of symbols x_f indexed by the nonconstant monic polynomials $f\in F[X]$. If 1 lies in the ideal I of $F[\dots,x_f,\dots]$ generated by the polynomials $f(x_f)$, then

$$g_1f_1(x_{f_1})+\cdots+g_nf_n(x_{f_n})=1\qquad(\text{in } F[\dots,x_f,\dots])$$

for some $g_i\in F[\dots,x_f,\dots]$ and some nonconstant monic $f_i\in F[X]$. Let E be an extension of F such that each f_i, $i=1,\dots,n$, has a root α_i in E. Under the F-homomorphism $F[\dots,x_f,\dots]\to E$ sending

$$\begin{cases} x_{f_i}\mapsto\alpha_i \\ x_f\mapsto 0, & f\notin\{f_1,\dots,f_n\}\end{cases}$$

the above relation becomes $0=1$. From this contradiction, we deduce that 1 does not lie in I, and so Proposition 6.4 applied to $F[\dots,x_f,\dots]/I$ shows that I is contained in a maximal ideal M of $F[\dots,x_f,\dots]$. Let $\Omega=F[\dots,x_f,\dots]/M$. Then Ω is a field containing (a copy of) F in which every nonconstant polynomial in $F[X]$ has at least one root. Repeat the process starting with E_1 instead of F to obtain a field E_2. Continue in this fashion to obtain a sequence of fields

$$F=E_0\subset E_1\subset E_2\subset\cdots,$$

and let $E=\bigcup_i E_i$. Then E is algebraically closed because the coefficients of any nonconstant polynomial g in $E[X]$ lie in E_i for some i, and so g has a root in E_{i+1}. Therefore, the algebraic closure of F in E is an algebraic closure of F .

(Non)uniqueness of algebraic closures

THEOREM 6.8 (*) *Let Ω be an algebraic closure of F, and let E be an algebraic extension of F. There exists an F-homomorphism $E \to \Omega$, and, if E is also an algebraic closure of F, then every such homomorphism is an isomorphism.*

PROOF. Suppose first that E is countably generated over F, i.e., $E = F[\alpha_1, ..., \alpha_n, \ldots]$. Then we can extend the inclusion map $F \to \Omega$ to $F[\alpha_1]$ (map α_1 to any root of its minimal polynomial in Ω), then to $F[\alpha_1, \alpha_2]$, and so on (see 2.2).

In the uncountable case, we use Zorn's lemma. Let S be the set of pairs (M, φ_M) with M a field $F \subset M \subset E$ and φ_M an F-homomorphism $M \to \Omega$. Write $(M, \varphi_M) \leq (N, \varphi_N)$ if $M \subset N$ and $\varphi_N | M = \varphi_M$. This makes S into a partially ordered set. Let T be a totally ordered subset of S. Then $M' = \bigcup_{M \in T} M$ is a subfield of E, and we can define a homomorphism $\varphi' : M' \to \Omega$ by requiring that $\varphi'(x) = \varphi_M(x)$ if $x \in M$. The pair (M', φ') is an upper bound for T in S. Hence Zorn's lemma gives us a maximal element (M, φ) in S. Suppose that $M \neq E$. Then there exists an element $\alpha \in E$, $\alpha \notin M$. Since α is algebraic over M, we can apply (2.2) to extend φ to $M[\alpha]$, contradicting the maximality of M. Hence $M = E$, and the proof of the first statement is complete.

If E is algebraically closed, then every polynomial $f \in F[X]$ splits in $E[X]$ and hence in $\varphi(E)[X]$. Let $\alpha \in \Omega$, and let $f(X)$ be the minimum polynomial of α. Then $X - \alpha$ is a factor of $f(X)$ in $\Omega[X]$, but, as we just observed, $f(X)$ splits in $\varphi(E)[X]$. Because of unique factorization, this implies that $\alpha \in \varphi(E)$. □

V.3 Exercises

1. Suppose $K = \mathbb{F}_q$ is a finite field of charakteristic p. Let $f = 1 + \prod_{\alpha \in K}(X - \alpha)$ be the polynomial used in the proof of V.1.3. Put $L = \Omega_K^f$. Show that:
   (a) $f = X^q - X + 1$;
   (b) for every $\alpha \in L$ such that $f(\alpha) = 0$ it holds that
   $$\alpha^{q^i} = \alpha - \bar{i} \quad \text{for all } i \in \mathbb{Z}_{>0};$$
   here $\bar{i} = (i \bmod p) \in \mathbb{F}_p \subset K$. In particular, show that
   $$\alpha^{q^p} = \alpha;$$
   (c) $L = \mathbb{F}_{q^p}$;
   (d) every irreducible factor of f in $K[X]$ has degree p.
2. Suppose $\overline{K}$ is an algebraic extension of the field K such that for every monic $f \in K[X]$ the field $\overline{K}$ contains a splitting field of f over K. Show that $\overline{K}$ is an algebraic closure of K.
3. Let $\overline{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Show that $[\overline{\mathbb{Q}} : \overline{\mathbb{Q}} \cap \mathbb{R}] = 2$.

VI MODULES

VI.1 Definitions

The analogue of a vectorspace over a field is a module over a ring.

VI.1.1 Definition. Let R be a (unitary) ring. A *left-R-module* M is an abelian group $(M,+,0)$ with an *action* of the ring R, which means a map:

$$R \times M \longrightarrow M, \qquad (a,m) \mapsto am,$$

such that for all $a, b \in R$ and all $m, n \in M$ it holds that:

(RM1) $a(m+n) = am + an$,
(RM2) $(a+b)m = am + bm$,
(RM3) $a(bm) = (ab)m$,
(RM4) $1m = m$.

A *right-R-module* is defined analogously, but in that case with an action $M \times R \to M$ et cetera.

VI.1.2 Remark. Unless explicitly stated otherwise, all modules considered here are left-R-modules and we simply write R-modules.

VI.1.3 Example. Let $R = K$ be a field. In this case the axioms for a left-R-module are exactly the axioms for a vectorspace over K. ∎

VI.1.4 Example. Let K be a field. The ring R consisting of all $n \times n$ matrices with coefficients in K acts on the additive group $(K^n)^+$ of the vectorspace K^n:

$$R = M(n,K), \quad M = (K^n)^+ \qquad \text{with action} \quad (A,v) \mapsto Av,$$

the usual product of a matrix and a vector. This turns K^n into a left-$M(n,K)$-module. ∎

VI.1.5 Example. Take $R = \mathbb{Z}$ and let $M = G$ be an abelian group. Define an action of $\mathbb{Z}$ on G using $0g := e$, $(-1)g := -g$ with $e \in G$ the unit element and $-g$ the inverse of g in G. Next, define $ng := g+g+\ldots+g$ (n times) and $(-n)g = (-g)+(-g)+\ldots+(-g)$ (n times) whenever $n \geq 1$. Verify that in this way G is a $\mathbb{Z}$-module. ∎

VI.1.6 Example. Suppose R is a ring and $I \subset R$ is an ideal. Then I is an R-module. Moreover R/I is an R-module as well, with action:

$$R \times R/I \longrightarrow R/I \qquad (r, a+I) \mapsto ra + I.$$

Note that in this way $r\overline{a} := ra + I = \overline{ra}$. ∎

VI.1.7 Remark. The axioms of a left-R-module imply:

$$0m = (0+0)m = 0m + 0m \qquad \text{hence} \quad 0m = 0.$$

Moreover:

$$0 = (a + (-a))m = am + (-a)m \qquad \text{so} \quad (-a)m = -(am),$$

in particular $(-1)m = -(1m) = -m$, hence $a(-m) = a((-1)m) = (a(-1))m = (-a)m$, which justifies the notation $-am$ for $(-a)m = -(am) = a(-m)$.

VI.1.8 Definition. A *submodule* N of a left-R-module M is a subgroup of M which is closed under the action of R. This means that $N \subset M$ has the property that for all $a, b \in N$ and all $r \in R$ one has:

(SM1) $0 \in N, \quad a - b \in N,$
(SM2) $ra \in N, \quad$ (one also writes $rN \subseteq N$).

A submodule N of the R-module M is itself also an R-module.

VI.1.9 Examples. 1. Take a field $R = K$, and let M be a vectorspace over K. The submodules of M are precisely the linear subspaces of M.
2. The only submodules of the $M(n,K)$-module K^n are $\{0\}$ and K^n. Namely, if $x, y \in K^n - \{0\}$ then $A \in M(n,K)$ exists with $Ax = y$ (verify this, and check how it implies the claim!).
3. The submodules of the $\mathbb{Z}$-module G, with G any abelian group, are precisely the subgroups of G (check for yourself!).
4. The ring R is itself an R-module, and the R-submodules of R are precisely the ideals of R (as one easily verifies!).
5. Take $A \in M(n,K)$ where K is a field, and let $R = K[A] = \{\sum_{i<\infty} a_i A^i : a_i \in K\}$. Suppose $v \in K^n$ is an eigenvector of A with eigenvalue $\lambda \in K$. Then the (one-dimensional) subspace Kv is a submodule of the R-module K^n. Namely, Kv is an additive group and since $Av = \lambda v \in Kv$, also $(\sum_{i<\infty} a_i A^i)(v) \in Kv$.

VI.2 R-module homomorphisms

The R-module homomorphisms generalise the linear maps:

VI.2.1 Definition. Let R be a ring and let M, N be (left)-R-modules. An *R-module homomorphism* is a map

$$f : M \longrightarrow N,$$

which is a homomorphism of abelian groups and which is moreover R-linear. In other words, for all $x, y \in M$ and for all $r \in R$ the map f satisfies:

(H1) $f(x+y) = f(x) + f(y),$
(H2) $f(rx) = rf(x).$

In particular it follows that $f(0) = 0$.
An *R-module isomorphism* is a bijective R-module homomorphism. (Note that the inverse of an R-module isomorphism is again a (bijective) R-module homomorphism and therefore it is again an R-module isomorphism.)
The *kernel* of an R-module homomorphism $f : M \to N$ is defined as

$$\mathrm{Ker}(f) := \{m \in M : \ f(m) = 0\,(\in N)\}.$$

The *image* of an R-module homomorphism is defined as

$$\mathrm{Im}(f) = f[M] = \{f(m) \in N : \ m \in M\}.$$

VI.2.2 Examples. 1. Since $0 = \{0\} \subset R$ is an R-module (it is an ideal in R), it follows that the obvious maps

$$0 \longrightarrow M, \qquad M \longrightarrow 0$$

are R-module homomorphisms (the first one sends $0 \in R$ to $0 \in M$, the second one sends every $m \in M$ to $0 \in R$).

2. If R is a commutative ring and M is an R-module and $a \in R$, then

$$\phi_a : M \longrightarrow M, \qquad m \mapsto am$$

is an R-module homomorphism.

3. A homomorphism $f : G \to H$ of abelian groups is a $\mathbb{Z}$-module homomorphism, namely $f(2g) = f(g+g) = f(g) + f(g) = 2f(g)$ et cetera.

4. Let R be a ring and $I \subset R$ an ideal. Then R and R/I are R-modules. The canonical map (a ring homomorphism):

$$\phi : R \longrightarrow R/I, \qquad m \mapsto \overline{m} = m + I,$$

is an R-module homomorphism. Namely, condition (H1) is satisfied since ϕ is a ring homomorphism and therefore a homomorphism of additive groups. Moreover (see Example VI.1.6):

$$\phi(rm) := rm + I = r\phi(m), \qquad (r, m \in R)$$

so also (H2) is satisfied.

VI.2.3 Theorem. *Let $f : M \to N$ be an R-module homomorphism. Then:*

1. $\mathrm{Ker}(f)$ *is a submodule of* M.
2. $\mathrm{Im}(f)$ *is a submodule of* N.

Proof. These assertions are immediate consequences of the definitions (check for yourself!). ∎

VI.2.4 Theorem. *If $f : M \to N$ and $g : N \to P$ are R-module homomorphisms, then the composition $g \circ f : M \to P$, so*

$$g \circ f : M \xrightarrow{f} N \xrightarrow{g} P$$

is an R-module homomorphism as well.

Proof. This is immediate from the definition. ∎

VI.2.5 Remark. One usually writes gf for the composition $g \circ f$ of the R-module homomorphisms f and g.

VI.3 Direct sums

All modules considered in this section are left-modules.

VI.3.1 Definition. Let R be a ring and let I be a nonempty set. For every $i \in I$ let M_i be an R-module. The *direct sum* M of the M_i's is defined as

$$M = \bigoplus_{i \in I} M_i := \left\{ (x_i)_{i \in I} : \ x_i \in M_i \text{ for all } i \in I \text{ and } x_i \neq 0 \text{ for only finitely many } i \right\}.$$

This set is given the structure of an R-module by defining

$$\begin{array}{rcll}(x_i)_{i\in I}+(y_i)_{i\in I} & = & (z_i)_{i\in I}, & \text{with } z_i = x_i + y_i \ \ \forall i \in I \\ 0 & = & (0)_{i\in I}, & (\text{so the } i\text{–th coordinate is } 0 \in M_i, \ \forall i \in I \\ r\cdot(x_i)_{i\in I} & = & (z_i)_{i\in I} & \text{with} \quad z_i = rx_i \ \ \forall i \in I.\end{array}$$

It is not hard to check that in this way indeed an R-module is defined.

VI.3.2 Definition. An R-module F is called *free* (or, free R-module) if a nonempty set I exists, and an isomorphism of R-modules

$$F \xrightarrow{\cong} \oplus_{i\in I} R.$$

(So here we take all modules M_i in VI.3.1 equal to R.)

Given a nonempty set I, the R-module $F := \oplus_{i\in I} R$ is by definition free. For every $i \in I$, define

$$e_i = (\ldots, x_j, \ldots)_{j\in I} \in F \qquad \text{by} \quad x_i = 1, \quad \text{and } x_j = 0 \text{ for all } j \neq i.$$

Then every $x \in F$ can be written in a unique way as

$$x = \sum_{i\in I} x_i e_i, \qquad (x, e_i \in F, \ x_i \in R)$$

(which is a finite sum since only finitely many $i \in I$ exist with $x_i \neq 0$).

As a special case, for $n \in \mathbb{Z}_{\geq 1}$ one defines the free R-module

$$R^n := \bigoplus_{i\in\{1,2,\ldots n\}} R.$$

VI.3.3 Examples. If K is a field, K^n is the familiar vector space over K.

The polynomial ring $R[X]$ is an R-module (for the usual addition and scalar multiplication). The map

$$\phi : R[X] \longrightarrow F := \bigoplus_{i\in\mathbb{Z}_{\geq 0}} R, \qquad \sum_{i=0}^{n} a_i X^i \mapsto (a_0, a_1, \ldots, a_n, 0, 0, \ldots),$$

is an isomorphism of R-modules; ϕ is surjective since only finitely many x_i differ from 0 in an element $(\ldots, x_i, \ldots) \in F$.

For R *non-commutative* rings R it turns out to be possible (see Exercise 7 on page 57) that $R \cong R^2$ (!). This is not possible in case R is commutative:

VI.3.4 Theorem. *If $R \neq (0)$ is a commutative (unitary) ring, then*

$$R^m \cong R^n \quad \Longrightarrow \quad m = n.$$

The number m $(= n)$ here is called the rank of the R-module R^m. More generally

$$R^m \cong \oplus_{i\in I} R \quad \Longrightarrow \quad m = \#I.$$

Proof. A proof based on the well known property $\det(AB) = \det(A)\det(B)$ of determinants of $n \times n$ matrices, is given in Section 3.4 of the book N. Jacobson, *Basic Algebra I* (San Fransisco: W.H. Freeman and Company, 1974). The advantage of that proof is that no use is made of Zorn's Lemma. The next argument uses the existence of a maximal ideal in a ring, which in general does require Zorn's Lemma.

Suppose $\phi : R^m \to R^n$ is an R-module isomorphism. Take M a maximal ideal in R. Then $M^m = \oplus_{i=1}^m M$ is a submodule of R^m. The factor group R^m/M^m is in a natural way a $K = R/M$-module, hence since K is a field, R^m/M^m is a vectorspace over K. One easily verifies that the map

$$(r_1,\ldots,r_m) + M^m \mapsto (r_1 + M,\ldots,r_m + M)$$

yields a well-defined isomorphism between vectorspaces R^m/M^m and K^m over K. Hence $\dim_K(R^m/M^m) = m$.

The set $\phi(M^m)$ is a submodule of R^n. The factor group $R^n/\phi(M^m)$ therefore is an R-module, and even a K-module: namely, for $m \in M$ and $v \in R^n$ one has $mv = \phi\phi^{-1}(mv) = \phi(m\phi^{-1}(v)) \in \phi(M^m)$. Hence multiplying an element $r \in R$ by a class $v + \phi(M^m) \in R^n/\phi(M^m)$ only depends on the class of r in $R/M = K$, which exactly means that $R^n/\phi(M^m)$ is a K-module. The isomorphism ϕ therefore induces an isomorphism $R^m/M^m \cong R^n/\phi(M^n)$ between vectorspaces over K, so in particular $\dim_K(R^n/\phi(M^m)) = m$.

Every $x \in R^n$ can be written in a unique way as

$$\begin{aligned} x &= (x_1, x_2, \ldots, x_n) \\ &= x_1 \cdot (1,0\ldots,0) + x_2 \cdot (0,1,0,\ldots,0) + \ldots + x_n \cdot (0,0,\ldots,1) \\ &= x_1 e_1 + x_2 e_2 + \ldots + x_n e_n. \end{aligned}$$

Hence the vectorspace $R^n/\phi(M^m)$ over K is spanned by $e_1 + \phi(M^m),\ldots,e_n + \phi(M^m)$. We conclude that $n \geq \dim_K(R^n/\phi(M^m)) = m$.

Interchanging the roles of R^n and R^m in the argument above, and replacing ϕ by ϕ^{-1}, it follows analogously that also $m \geq n$. Hence $n = m$.

The more general assertion in the theorem follows completely analogously. This shows Theorem VI.3.4. ∎

We saw that modules over a field K are in fact vectorspaces over K. In particular: if K^n is the direct sum of two K-modules:

$$K^n \cong V \oplus W, \quad \text{then} \quad V \cong K^a \quad \text{and} \quad W \cong K^{n-a}$$

for some a, since V and W are (finite dimensional) vectorspaces over K as well. The case of modules over a ring that is not a field is much more interesting. For a simple example, take the ring $R = \mathbb{Z}/6\mathbb{Z}$ and the R-modules $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$, then

$$\mathbb{Z}/6\mathbb{Z} \xrightarrow{\cong} \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}, \qquad n + 6\mathbb{Z} \mapsto (n + 2\mathbb{Z}, n + 3\mathbb{Z}).$$

Using the Chinese Remainder Theorem many more examples of this kind can be constructed. The next two examples reveal a slightly more surprising submodules of a free module over a ring.

VI.3.5 Example. Take $R = \mathbb{Z}[\sqrt{-5}]$. The map $\varphi\colon R \to \mathbb{F}_2$ defined by $\varphi(a+b\sqrt{5}) = \overline{a}+\overline{b}$, with $a,b \in \mathbb{Z}$ and for any $n \in \mathbb{Z}$ using the notation $\overline{n} := n \bmod 2$, is the surjective ring homomorphism. Hence $M := \mathrm{Ker}(\varphi)$ is a maximal ideal in R, so in particular M is an R-module. Note that

$$\begin{aligned} M &= \left\{a + b\sqrt{-5} : \exists k \in \mathbb{Z} \text{ such that } a + b = 2k\right\} \\ &= \left\{2k - b + b\sqrt{-5} : b,k \in \mathbb{Z}\right\} \\ &= \mathbb{Z} \cdot 2 + \mathbb{Z} \cdot (-1 + \sqrt{-5}) \end{aligned}$$

and every element in M can be written in a unique way as $n \cdot 2 + m \cdot (-1 + \sqrt{-5})$ with $n,m \in \mathbb{Z}$.

We claim that

$$M \oplus M \cong R^2$$

as R-modules. Indeed, the map $(\alpha,\beta) \mapsto \big(2\alpha+(-1+\sqrt{-5})\beta, (1+\sqrt{-5})\alpha - 2\beta\big)$ for $\alpha,\beta \in R^2$ defines an R-module homomorphism from R^2 to $M \oplus M$. The "inverse" R-module homomorphism $M \oplus M \to R^2$ is given by

$$\begin{array}{c} \big(a\cdot 2 + b\cdot(-1+\sqrt{-5}), c\cdot 2 + d\cdot(-1+\sqrt{-5})\big) \\ \mapsto \\ \big(-2a+b+c+2d+(d-b-c)\sqrt{-5}, -a+3b+2c-d+(d-a)\sqrt{-5}\big) \end{array}$$

for $a,b,c,d \in \mathbb{Z}$.

Next, we claim that M itself is *not* free as an R-module. Indeed, if $M \cong R^n$ for some n, then $R^2 \cong M \oplus M \cong R^n \oplus R^n \cong R^{2n}$. So we necessarily have $n = 1$. However, if $M \cong R$ as R-modules, then an isomorphism $f : R \to M$ would send $1 \in R$ to some element $e = f(1) \in M$, and then any $r \in R$ to $f(r) = f(r\cdot 1) = r \cdot f(1) = re \in M$. As f is supposed to be an isomorphism, this would imply $M = Re$, so M is a principal ideal. In particular, the norm of any element of M would be a multiple of the norm of e. Now $2 \in M$ has norm 4 and $-1+\sqrt{-5}$ has norm 6, so the norm of e divides $\gcd(4,6) = 2$. Since $R = \mathbb{Z}[\sqrt{-5}]$ does not contain any element of norm 2, we conclude that e must have norm 1, so $e \in R^\times$ and therefore $M = Re = R$, which is absurd since, e.g., $1 \notin M$.

So in this example the module M is not free, while $M \oplus M$ is free! ▬

VI.3.6 Example. Consider the subring R of periodic functions with period 2π of the ring of all real C^∞-functions on $\mathbb{R}$:

$$R := \big\{ f \in C^\infty(\mathbb{R}) :\ f(x+2\pi) = f(x)\ \ \forall\, x \in \mathbb{R} \big\}.$$

Let M be the R-module defined as

$$M := \big\{ m \in C^\infty(\mathbb{R}) :\ m(x+2\pi) = -m(x)\ \ \forall\, x \in \mathbb{R} \big\}.$$

Here addition and scalar multiplication on M are given by

$$\begin{array}{rcll} (m+n)(x) & := & m(x)+n(x), & m,n \in M,\ x \in \mathbb{R}, \\ (fm)(x) & := & f(x)m(x), & f \in R,\ m \in M,\ x \in \mathbb{R}; \end{array}$$

note that indeed $fm \in M$ (!) and check for yourself that in this way M is an R-module. We claim that like in Example VI.3.5 we have that

$$M \ncong R, \qquad \text{whereas} \quad M \oplus M \cong R^2,$$

and M is not free.

Well known properties of the sine and cosine functions imply

$$\forall\, a \in \mathbb{R} :\ C_a, S_a \in M, \quad \text{where} \quad \begin{cases} C_a : \mathbb{R} \to \mathbb{R}, & x \mapsto \cos\frac{x-a}{2}, \\ S_a : \mathbb{R} \to \mathbb{R}, & x \mapsto \sin\frac{x-a}{2}. \end{cases}$$

Suppose an R-module isomorphism between R and M exists:

$$\phi : R \longrightarrow M, \qquad \text{then} \quad \phi(f) = f \cdot \phi(1)$$

with $1 \in R$ the constant function 1. Write $g := \phi(1) \in M$. Since g is continuous and $g(2\pi) = -g(0)$ one concludes that g has a zero $a \in [0, 2\pi]$.

As ϕ is supposed to be surjective, every $m \in M$ can be written as $m = \phi(f)$ for some $f \in R$, hence

$$m = fg, \qquad \text{implying} \quad m(a) = f(a)g(a) = 0.$$

However, taking $m = C_a \in M$ one obtains $C_a(a) = \cos\frac{a-a}{2} = 1 \neq 0$, a contradiction. We conclude that no isomorphism $\phi : R \to M$ exists, which shows the first assertion.

Next we show $R^2 \cong M \oplus M$. Define

$$\psi : R^2 \longrightarrow M \oplus M, \qquad (f, g) \mapsto (fC_0 + gS_0,\, -fS_0 + gC_0)$$

(this clearly is an R-module homomorphism). In terms of matrices (with coefficients in $M \subset C^\infty(\mathbb{R})$) this ψ is given by

$$A := \begin{pmatrix} C_0 & S_0 \\ -S_0 & C_0 \end{pmatrix}, \qquad \text{so in particular} \quad A^{-1} = \begin{pmatrix} C_0 & -S_0 \\ S_0 & C_0 \end{pmatrix},$$

where we used that $C_0^2 + S_0^2 = 1$.

We claim that indeed the inverse of ψ is given by

$$\psi^{-1} \colon M \oplus M \longrightarrow R^2, \qquad (m, n) \mapsto (C_0 m - S_0 n,\, S_0 m + C_0 n).$$

To verify this, first note that the product of any two functions in M is an element of R, hence ψ^{-1} indeed maps $M \oplus M$ to R^2). Clearly ψ^{-1} is an R-module homomorphism, and $\psi^{-1}\psi = \mathrm{id}_{R^2}$ and $\psi\psi^{-1} = \mathrm{id}_{M\oplus M}$. So ψ^{-1} is the inverse of ψ.

Finally, we show that M is not free. Exactly as in Example VI.3.5 above, if $M \cong \oplus_{i\in I} R$ for some set I, then $\#I = 1$ and $R \cong M$. Since we already showed that $R \ncong M$ it follows that M is not free. ∎

VI.3.7 Remark. The module M in Example VI.3.6 turns out to be isomorphic (as an R-module) to some (non-principal) ideal in R (see Exercise 3 on page 56):

$$M \cong I := \mathrm{Ker}(\mathrm{ev}_0 \colon R \longrightarrow \mathbb{R}), \qquad \text{with} \quad \mathrm{ev}_0 \colon f \mapsto f(0).$$

The next result describes a "universal property" of direct sums: this means a property determining it up to isomorphisms of modules.

VI.3.8 Theorem. *Let R be a ring and let I be a (nonempty) set. Suppose that for every $i \in I$ an R-module M_i is given. Then $\oplus_{i\in I} M_i$ together with the R-module homomorphisms $\iota_i : M_i \to \oplus_{i\in I} M_i$ defined for $m \in M_i$ by $\iota_i(m) = (x_j)_{j\in I}$ (where $x_j = 0 \in M_j$ if $j \neq i$ and $x_i = m$), have the following property.*

> *For every R-module M, given R-module homomorphisms $f_i : M_i \to M$ for all $i \in I$, there exists a unique(!) R-module homomorphism $f : \oplus_{i\in I} M_i \to M$ such that $f_i = f \circ \iota_i$ for every $i \in I$.*

Is D any R-module equipped with homomorphisms $j_i : M_i \to D$ such that the same property holds:

> *For every R-module M, given R-module homomorphisms $f_i : M_i \to M$ for all $i \in I$, there exists a unique(!) R-module homomorphism $f : D \to M$ such that $f_i = f \circ j_i$ for every $i \in I$,*

then $D \cong \oplus_{i\in I} M_i$.

Proof. Given maps $f_i : M_i \to M$, define $f : \oplus_{i\in I} M_i \to M$ by $f((x_i)_{i\in I}) = \sum_i f_i(x_i)$. By the definition of 'direct sum' and the fact that any R-module homomorphism sends 0 to 0, this is a finite sum. Evidently f is an R-module homomorphism, and it has the property $f_i = f \circ \iota_i$ for all $i \in I$. Since any $(m_i)_{i\in I} \in \oplus_{i\in I} M_i$ can be written as a finite sum $\sum_{i\in I} \iota_i(m_i)$, it follows that any R-module homomorphism $\oplus_{i\in I} M_i \to M$ is completely determined by its restrictions to the $\iota_i(M_i)$, $i \in I$. So f is unique.

If D, with the maps j_i, has the same property, then we apply this property to the R-module $\oplus_{i\in I} M_i$ and the homomorphisms ι_i. This gives us a unique $f : D \to \oplus_{i\in I} M_i$ such that $\iota_i = f \circ j_i$ for all i.

Reversing the roles of D and $\oplus_{i\in I}M_i$ one obtains a unique $g\colon \oplus_{i\in I}M_i \to D$ with $j_i = g\circ\iota_i$. Combining the two conclusions yields $\iota_i = (f\circ g)\circ\iota_i$ which, by applying the 'universal property' of $\oplus_i M_i$ and the maps ι_i, shows that $f\circ g$ is the identity map on $\oplus_i M_i$.

Finally, we also have $j_i = (g\circ f)\circ j_i$, which by the uniqueness in the property we assume for D implies that $g\circ f$ is the identity on D. So indeed D and $\oplus_i M_i$ are isomorphic, finishing the proof. ∎

VI.4 Cyclic modules

In this section we discuss modules M over a ring R, such that we have $M = Re$ for some $e \in M$. It turns out that this simple concept has some surprising applications.

VI.4.1 Definition. For a ring R, a (left) R-module M is called *cyclic* if $e \in M$ exists with $M = Re$.

VI.4.2 Example. An abelian group, regarded as a module over $\mathbb{Z}$, is cyclic as a group if and only if it is cyclic as a $\mathbb{Z}$-module. Since any cyclic group is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ for some $n \geq 0$ (the case $n = 0$ is exactly the case of an infinite cyclic group!), we have that the cyclic $\mathbb{Z}$-modules are, up to isomorphisms, the modules $\mathbb{Z}/n\mathbb{Z}$ with $n \geq 0$. We will generalize this example in Theorem VI.4.4 below. ▁∎

VI.4.3 Example. Suppose the ring R is *commutative* and let $I \subset R$ be an ideal. Then I is an R-module, and it is a cyclic R-module if and only if the ideal I is a principal ideal. ▁∎

VI.4.4 Theorem. *If R is a commutative ring (with 1) and M is a cyclic R-module, then $I := \{r \in R : rm = 0\ \forall m \in M\}$ is an ideal in R and $M \cong R/I$ as R-modules.*

We have that M is free (and in that case of rank 1) precisely when $I = (0)$.

Proof. Write $M = Re$ for some $e \in M$. The map $R \to M$ given by $r \mapsto re$ is an R-module homomorphism, with kernel exactly I. Hence I is an R-module, and since $I \subset R$ this means it is an ideal in R. As the given map is surjective, one concludes $R/I \cong M$ as desired.

If $I = (0)$ then $R \cong M$ as R-modules, so indeed M is free (of rank 1). And if M is free, say $M \cong \oplus_{i\in S}R$ for some nonempty set S, then

$$I = \{r \in R : rM = (0)\} = \{r \in R :\ \cdot(r \oplus_{i\in S} R) = (0)\} = (0).$$

This finishes the proof. ∎

VI.4.5 Example. Let K be a field and take V a finite dimensional vectorspace over K. Suppose $\varphi\colon V \to V$ is a K-linear map. Evaluating at φ defines a ring homomorphism from the polynomial ring $K[X]$ to the ring $\mathrm{End}(V)$ of all K-linear maps $V \to V$. The image of this map is denoted $R := K[\varphi]$; it is a commutative subring of $\mathrm{End}(V)$ and it consists of all linear maps $\sum_{j=0}^{n} a_j\varphi^j$, for $n \in \mathbb{Z}_{\geq 0}$ and all $a_j \in K$, here φ^0 denotes the identity map on V.

The vectorspace V obtains the structure of a $K[X]$-module by defining

$$(\sum_{j=0}^{n} a_j X^j)(v) = \sum_{j=0}^{n} a_j\varphi^j(v).$$

Given any $v \in V$ we have the cyclic submodule $K[X]v \subset V$. By construction, it is the K-linear subspace of V spanned by $v, \varphi(v), \varphi^2(v), \ldots, \varphi^n(v), \ldots$.

The vectorspace (cyclic $K[X]$-module) $K[X]v$ is called the *Krylov subspace* (corresponding to φ and v), named after the Russian naval engineer and mathematician Aleksey Nikolaevich Krylov (1863–1945).

Using Theorem VI.4.4 we see that $K[X]v \cong K[X]/I$ with $I \in K[X]$ the ideal consisting of all $\sum_{j=0}^{n} a_j X^j$ such that $\sum_{j=0}^{n} a_j \varphi^j(v) = 0$. As $K[X]$ is a principal ideal domain, $I = (g)$ for some polynomial g. Clearly $g \neq 0$ (the vectorspace V was assumed to be finite dimensional, so there exist certainly linear dependencies between the vectors $\varphi^j(v)$!). Hence we may and will assume that g is monic. Certainly the kernel of the evaluation map $K[X] \to K[\varphi]$ is contained in $I = (g)$. The latter kernel is the principal ideal generated by the minimal polynomial m_φ of φ. So $(m_\varphi) \subset (g)$, which means that g is a divisor of m_φ.

In particular

$$K[x]v \cong K[x]/(g)$$

and $\dim_K(K[x]v) = \deg(g) \leq \deg(m_\varphi)$. So for V to be cyclic as a $K[X]$-module, a necessary condition is that $\deg(m_\varphi) = \dim_K(V)$. Theorem VI.5.7 will show that this condition also suffices. ∎

VI.4.6 Example. *(here I intend to discuss the proof of N. Katz for the fact that a $K(t)\langle \frac{d}{dt} \rangle$-module that is finite dimensional as $K(t)$-module, is cyclic. As a consequence, given an $n \times n$ system of first order linear differential equations we can find a corresponding scalar linear differential equation of order n ...)* ∎

VI.5 An upper triangular form for matrices

Let K be a field and let V be a finite dimensional vectorspace over K, and let $\alpha : V \to V$ be a K-linear map. Under the condition that all eigenvalues of α are in K, we will construct a basis of V such that the matrix $A = (a_{ij})$ of α with respect to this basis is upper triangular (this means $a_{ij} = 0$ for $i > j$).

With α as above, as in Example VI.4.5 let

$$K[\alpha] := \mathrm{ev}_\alpha(K[X]) = \{ \sum_{i<\infty} a_i \alpha^i : \ a_i \in K \}$$

be the image of the evaluation homomorphism

$$\mathrm{ev}_\alpha : K[X] \longrightarrow \mathrm{End}_K(V), \qquad f \mapsto f(\alpha).$$

We have that $K[\alpha] \cong K[X]/(m_\alpha)$ with $m_\alpha \in K[X]$ the minimal polynomial of α. Note that m_α is a divisor of the characteristic polynomial $P_\alpha = \det(\alpha - XI)$, and every irreducible factor of P_α is also an irreducible factor of m_α. Nevertheless, there are many examples with $\deg(m_\alpha) < \deg(P_\alpha)$.

The vectorspace V is a $K[\alpha]$-module with

$$K[\alpha] \times V \longrightarrow V, \qquad (\sum_{i=0}^{n} a_i \alpha^i, v) \mapsto \sum_{i=0}^{n} a_i \alpha^i(v).$$

Since $K \subset K[\alpha]$, any $K[\alpha]$-submodule of V is also a K-vectorspace. We will write V as a direct sum of $K[\alpha]$-submodules of V. In each of these submodules the action of α will have a simple description.

The principal ideal domain $K[X]$ is a unique factorization domain, hence we can write

$$m_\alpha = h_1^{n_1} h_2^{n_2} \ldots h_k^{n_k},$$

with all h_i monic and irreducible in $K[X]$ and $h_i \neq h_j$ for $i \neq j$. This factorization is unique up to permuting the indices $1, 2, \ldots k$.

We will use this factorization to construct a direct sum decomposition of the $K[\alpha]$-module V.

VI.5.1 Theorem. *Let K be a field and let $V \neq (0)$ be a finite dimensional vectorspace over K en let*

$$\alpha : V \longrightarrow V$$

be a linear map, with minimal polynomial

$$m_\alpha = h_1^{n_1} h_2^{n_2} \ldots h_k^{n_k}$$

for irreducible monic $h_i \in K[X]$ and $h_i \neq h_j$ for $i \neq j$.

Put

$$V_i := \left\{ v \in V \ : \ h_i^{n_i}(\alpha) v = 0 \right\}.$$

Then each V_i is a $K[\alpha]$-module. Moreover $V_i \neq 0$ for $i = 1, 2, \ldots, k$ and

$$V \cong \bigoplus_{i \in \{1,2,\ldots,k\}} V_i.$$

Proof. The fact that each V_i is a $K[\alpha]$-module is immediate from the definitions. For the remaining assertions we use mathematical induction with respect to k. If $k = 1$ then $h_1^{n_1}(\alpha) = m_\alpha(\alpha) = 0$ hence $V = V_1$, from which the result follows for $k = 1$.

Now let $k > 1$. Define

$$h := h_k^{n_k}, \qquad f := h_1^{n_1} h_2^{n_2} \ldots h_{k-1}^{n_{k-1}}, \quad \text{and} \quad N := \mathrm{Ker}(f(\alpha) \colon V \to V).$$

Since h and f are coprime, there exist $g_1, g_2 \in K[X]$ with

$$g_1 f + g_2 h = 1, \quad \text{so in particular} \quad g_1(\alpha) f(\alpha) + g_2(\alpha) h(\alpha) = 1.$$

We claim that the $K[\alpha]$-module homomorphism

$$\psi \colon N \oplus V_k \longrightarrow V \qquad (n, v_k) \mapsto n + v_k$$

is a $K[\alpha]$-module isomorphism.

Indeed, every $v \in V$ can be written as

$$v = 1v = f(\alpha) g_1(\alpha) v + h(\alpha) g_2(\alpha) v.$$

Now put $v_k = f(\alpha) g_1(\alpha) v$ and $n = h(\alpha) g_2(\alpha) v$, so $v = n + v_k$. As $m_\alpha(\alpha) = f(\alpha) h(\alpha) = 0$ one finds

$$\begin{aligned} h(\alpha) v_k &= m_\alpha(\alpha) g_1(\alpha) v = 0 \quad \text{so } v_k \in V_k, \text{ and} \\ f(\alpha) n &= m_\alpha(\alpha) g_2(\alpha) v = 0 \quad \text{so } n \in N. \end{aligned}$$

This shows that ψ is surjective. Moreover for $n \in N$ en $v_k \in V_k$ we have

$$n + v_k = 0 \implies x := n = -v_k \in N \cap V_k.$$

Therefore $f(\alpha) x = 0$ as well as $h(\alpha) x = 0$. From $x = 1x$ and $1 = f(\alpha) g_1(\alpha) + h(\alpha) g_2(\alpha)$ now follows that $x = 0$. So $\mathrm{Ker}(\psi) = (0)$ which shows ψ is injective, and we conclude that $V \cong N \oplus V_k$.

Let β be the restriction of α to N. Then it is easy to verify that $\beta : N \to N$. We have (compare Exercise 9 on page 58) that $m_\beta = f$. The induction hypothesis therefore implies

$$N \cong \oplus_{i=1}^{k-1} V_i, \qquad \text{hence we conclude} \quad V \cong N \oplus V_k \cong V_1 \oplus V_2 \oplus \ldots V_k.$$

Finally $V_i \neq \{0\}$ since otherwise $h_i(\alpha)^{n_i} : V \to V$ would be injective, and therefore (since $\dim_K(V) < \infty$) invertible. Then $m_\alpha(\alpha) = 0$ implies $m_\alpha(\alpha) h_i(\alpha)^{-n_i} = 0$, hence $m_\alpha h_i^{-n_i}$ is a polynomial in the kernel of the evaluation homomorphism ev_α, and it has lower degree than m_α. This contradiction finishes the proof. ∎

VI.5.2 Remark. The V_i in Theorem VI.5.1 are called the *generalized eigenspaces* of the linear map α.

VI.5.3 Example. Suppose $\alpha : K^n \to K^n$ is a linear map having n pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_n \in K$. In this case

$$m_\alpha = (X - \lambda_1) \ldots (X - \lambda_n)$$

and

$$V_i = \mathrm{Ker}(\alpha - \lambda_i), \qquad \text{so} \quad V_i := \{ v \in V : \ \alpha v = \lambda_i v \},$$

which means V_i is the eigenspace of α at the eigenvalue λ_i.

As each $V_i \neq \{0\}$ and $\sum_i \dim_K V_i = n$, every V_i has dimension one. Choose for $i = 1, 2, \ldots, n$ an $f_i \in V_i - \{0\}$, so f_i is an eigenvector of α with eigenvalue λ_i. Then $V_i = K \cdot f_i$ and

$$V = \oplus_{i=1}^n V_i = K \cdot f_1 \oplus \ldots \oplus K \cdot f_n$$

implies that the f_i are independent over K. Hence the f_i form a basis of K^n. Since $\alpha f_i = \lambda_i f_i$ for every i, the matrix of α with respect to this basis is the diagonal matrix $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. ∎

In the remainder of this section the special case where

$$m_\alpha = (X - \lambda_1)^{n_1} (X - \lambda_2)^{n_2} \ldots (X - \lambda_k)^{n_k},$$

with the $\lambda_i \in K$ pairwise distinct, is considered.

In particular the λ_i are exactly the eigenvalues of α. By Theorem VI.5.1 we have $V = \oplus_{i=1}^k V_{\lambda_i}$ as $K[\alpha]$-modules, with

$$V_{\lambda_i} := \mathrm{Ker}((\alpha - \lambda_i)^{n_i}) \neq \{0\}.$$

The fact that V_{λ_i} is a $K[\lambda]$-module implies in particular that if $v_i \in V_{\lambda_i}$ then also $\alpha(v_i) \in V_{\lambda_i}$. This means that a matrix of α with respect to a suitable basis will consist of blocks:

$$V = \oplus_{i=1}^k V_{\lambda_i} \xrightarrow{\alpha} \oplus_{i=1}^k V_{\lambda_i}, \qquad \alpha(V_{\lambda_i}) \subset V_{\lambda_i}.$$

We consider each of these blocks separately, in other words we restrict α to V_{λ_i}. Note that the restriction of α to V_{λ_i} has minimal polynomial $(X - \lambda_i)^{n_i}$.

VI.5.4 Example. Take $V = \mathbb{R}^3$ and

$$A := \begin{pmatrix} 4 & -4 & 4 \\ 1 & -1 & 4 \\ 0 & -1 & 4 \end{pmatrix}.$$

We determine the minimal polynomial of A and we find a basis of V on which A is given by blocks. First compute the characteristic polynomial of A, so

$$\det(A - XI) = -(X^3 - 7X^2 + 16X - 12) = -(X - 2)^2 (X - 3).$$

The minimal polynomial of A is therefore either $(X - 2)(X - 3)$ or $(X - 2)^2 (X - 3)$. One checks that $(A - 2)(A - 3) \neq 0$, and hence

$$m_A = (X - 2)^2 (X - 3).$$

The generalized eigenspaces are the kernels of

$$(A - 2)^2 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix} \quad \text{and of} \quad A - 3 = \begin{pmatrix} 1 & -4 & 4 \\ 1 & -4 & 4 \\ 0 & -1 & 1 \end{pmatrix}.$$

Hence (writing V_λ for the generalized eigenspace at the eigenvalue λ) one computes

$$V_2 = K\cdot(1,1,0)+K\cdot(1,1,1), \quad \text{and} \quad V_3 = K\cdot(0,1,1).$$

Since

$$\left.\begin{aligned} A(1,1,0) &= (0,0,-1) = 1\cdot(1,1,0)+(-1)\cdot(1,1,1) \\ A(1,1,1) &= (4,4,3) = 1\cdot(1,1,0)+3\cdot(1,1,1) \end{aligned}\right\},$$

we find that $\left(\begin{smallmatrix} 1 & 1 \\ -1 & 3 \end{smallmatrix}\right)$ is the matrix of the restriction of A to V_2 (with respect to the basis $(1,1,0)$, $(1,1,1)$ of V_2). Restricting A to V_3 yields the 1×1 matrix (3). The matrix of A in terms of the basis of V given by $\{(1,1,0),(1,1,1),(0,1,1)\}$ is therefore

$$\begin{pmatrix} 1 & 1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

■

VI.5.5 Theorem. *Let W be a finite dimensional vectorspace over a field K and let $B : W \longrightarrow W$ be a linear map with minimal polynomial $m_B = (X-\lambda)^m$.*

Then

$$B = \lambda I + N, \qquad with \quad N^m = 0$$

(here I denotes the identity map).
There exists a basis of W such that the matrix of B with respect to that basis has the form

$$\begin{pmatrix} \lambda & * & \dots & \dots & * \\ 0 & \lambda & * & \dots & * \\ 0 & 0 & & & * \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}.$$

Proof. The first statement follows by writing $N := B-\lambda I$ and noting that $m_B(B) = 0$.

Next, for $i = 1,2,\dots m$ define the linear subspace $W_i \subset W$ by

$$W_1 = \mathrm{Ker}(N),\dots,W_i = \mathrm{Ker}(N^i),\dots,W_m = \mathrm{Ker}(N^m) = \mathrm{Ker}(0) = W.$$

Note that $W_i \subset W_{i+1}$, since if $N^i v = 0$ then also $N^{i+1}v = 0$. Now take a basis of W starting from a basis of W_1, extending this to a basis of W_2, and continue in this way by extending a basis of W_i to a basis of W_{i+1}. At last this yields a basis $\{f_i\}$ of $W = W_m$.

Note that

$$NW_i \subset W_{i-1}, \qquad \text{because if} \quad N^i v = 0 \text{ then } N^{i-1}(Nv) = 0, \text{ so } Nv \in W_{i-1}.$$

Now let f_j be any of the chosen basis vectors. If $f_j \in W_1$ then $N(f_j) = 0$. If $f_j \notin W_1$, then i exists such that

$$f_j \in W_i \setminus W_{i-1}.$$

Since $NW_i \subset W_{i-1}$ we have by the construction of the basis that

$$N(f_j) = \sum_k x_{kj} f_k \qquad \text{with} \quad x_{kj} = 0 \text{ for } k \geq j.$$

Hence the matrix of N with respect to the basis $\{f_k\}$ is an upper triangular matrix with zeros on the diagonal. Hence $B = \lambda I + N$ is on the same basis given by the asserted form. This completes the proof. ■

VI.5.6 Example. Consider the $\mathbb{R}$-linear map

$$B\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \qquad \text{with matrix} \quad \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}.$$

The minimal polynomial m_B is a divisor of the characteristic polynomial

$$\det(B - XI) = X^2 - 4X + 4 = (X-2)^2$$

of B. Since $B - 2I \neq 0$, it follows that

$$m_B = (X-2)^2, \qquad \text{and} \quad N := B - 2I = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Check, as predicted by Theorem VI.5.5, that indeed $N^2 = 0$. We determine the W_i's:

$$W_1 = \mathrm{Ker}(N) = \mathbb{R}\cdot(1,1), \quad W_2 = \mathbb{R}^2.$$

As basis of $\mathbb{R}^2$ we take $f_1 := (1,1) \in W_1$, supplemented with $f_2 := (0,1) \in W_2$. Then

$$Nf_1 = 0, \quad Nf_2 = \tbinom{1}{1} = f_1.$$

Therefore $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ is the matrix of N with respect to the basis f_1, f_2. On the same basis the matrix van B has the upper triangular form

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

■

The arguments used in this section also make it possible to give a criterion for a K-vectorspace to be cyclic as a $K[\varphi]$-module, where $\varphi\colon V \to V$ is a linear map. This extends the discussion we started in Example VI.4.5. The result is the following.

VI.5.7 Theorem. *Let $V \neq (0)$ be a finite dimensional vectorspace over a field K and let $\varphi\colon V \to V$ be a linear map with minimal polynomial m_φ.*

Then V is cyclic as a $K[\varphi]$-module if and only if $\deg(m_\varphi) = \dim_K(V)$.

Proof. If $V = K[\varphi]\cdot v$ for some $v \in V$, then (as also discussed in Example VI.4.5) by Theorem VI.4.4 we have $V \cong K[\varphi]/I$ as $K[\varphi]$-modules, for some ideal $I \subset K[\varphi]$. Since as rings $K[\varphi] \cong K[X]/(m_\varphi)$, the ideal I corresponds to an ideal $J \subset K[X]$ with $(m_\varphi) \subset J$ and $K[X]/J \cong K[\varphi]/I \cong V$. As $K[X]$ is a principal ideal domain, $J = (g)$ for some $g \in K[X]$ and the property $(m_\varphi) \subset J$ means that $g | m_\varphi$. Now

$$\dim_K(V) = \dim_K(K[X]/(g)) = \deg(g)$$

implies that $\deg(m_\varphi) \geq \dim_K(V)$. Since m_φ divides the characteristic polynomial of φ which is of degree $\dim_K(V)$, we also have $\deg(m_\varphi) \leq \dim_K(V)$, hence it follows that $\deg(m_\varphi) = \dim_K(V)$.

For the converse, assume that $\deg(m_\varphi) = \dim_K(V)$. Factor

$$m_\varphi = h_1^{n_1} h_2^{n_2} \cdot \ldots \cdot h_k^{n_k}$$

for irreductible monic and pairwise distinct $h_i \in K[X]$, and all $n_i > 0$. Then by Theorem VI.5.1 $V \cong \bigoplus_i V_i$ as $K[\varphi]$-modules, with $V_i = \mathrm{Ker}(h_i^{n_i}(\varphi))$.

From the proof of VI.5.1 we know that the restriction of φ to V_i has minimal polynomial $h_i^{n_i}$. Hence $v_i \in V_i$ exists with $h_i^{n_i-1}(v_i) \neq 0$. We claim that

$$v := v_1 + v_2 + \ldots + v_k \in V$$

has the property $K[\varphi]\cdot v = V$. Indeed, suppose $f(\varphi)(v) = 0$ for some $f \in K[X]$. Then $f(\varphi)(v_1) = -f(\varphi)(v_2 + \ldots + v_k)$. As the left-hand side is in V_1 and the right-hand side in $V_2 \oplus \ldots \oplus V_k$, the fact that V is the direct sum of the subspaces V_i implies $f(\varphi)(v_1) = 0$ and $f(\varphi)(v_2 + \ldots + v_k) = 0$. Continuing inductively we find $f(\varphi)(v_i) = 0$ for all i. This implies that $h_i^{n_i} | f$, and therefore $m_\varphi | f$.

As $m_\varphi(\varphi) = 0$ one concludes that the ideal $I \subset K[\varphi]$ corresponding via Theorem VI.4.4 to the cyclic module $K[\varphi]\cdot v$, is the ideal (0). Hence $K[\varphi]\cdot v \cong K[\varphi]$. The latter K-vectorspace has dimension $\dim_K(V)$ by assumption, so we conclude that $K[\varphi]\cdot v \subset V$ has dimension $\dim_K(V)$ as well. This implies $K[\varphi]\cdot v = V$ hence V is cyclic as a $K[\varphi]$-module, finishing the proof. ∎

VI.6 Exercises

1. Take $R = \mathbb{R}[X]$. For $a \in \mathbb{R}$ define

$$I_a := (X-a) = \mathbb{R}[X] \cdot (X-a)$$

which is an ideal of R (and therefore an R-module).

   (a) Another R-module is the field of rational functions in the variable X over R. For $a, b \in \mathbb{R}$ define

$$\phi : I_a \longrightarrow \mathbb{R}(X) \quad \text{by} \quad f \mapsto f \cdot \frac{X-b}{X-a}.$$

   Verify that ϕ is an injective R-module homomorphism and that $\mathrm{Im}(\phi) = I_b$. Conclude that I_a and I_b are isomorphic as R-modules.

   (b) Show that the R-modules $\mathbb{R}_a := R/I_a$ and $\mathbb{R}_b := R/I_b$ are *not* isomorphic in case $a \neq b$.

2. Find using the methods of this chapter an upper triangular form for the following matrices:

$$A = \begin{pmatrix} 3 & -1 & 0 \\ 4 & -2 & 1 \\ 4 & -4 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ -1 & -2 & 3 \\ -1 & -3 & 4 \end{pmatrix}.$$

3. Let R be the ring of C^∞ functions $\mathbb{R} \to \mathbb{R}$ with period 2π, and

$$M := \{ g \in C^\infty(\mathbb{R}) : \, g(x+2\pi) = -g(x) \}, \quad \text{and} \quad I := \{ f \in R : \, f(0) = 0 \}.$$

Then both M and I are R-modules (for M see Example VI.3.6, and $I \subset R$ is an ideal).

   (a) Show that

$$\phi : M \longrightarrow I, \qquad g \mapsto gS_0,$$

   with $S_0(x) := \sin \frac{x}{2}$, is an injective R-module homomorphism.

   (b) Take $f \in I$. Show that the function g defined by

$$g(x) := \begin{cases} f(x)/S_0(x), & \text{if} \quad x \not\equiv 0 \bmod 2\pi, \\ 2f'(x) & \text{if} \quad x \equiv 0 \bmod 2\pi \end{cases}$$

   is a C^∞ function. Hint: check that $g \in M$ using

$$f(x) = \int_0^1 \frac{\partial f}{\partial t}(tx)dt = x \int_0^1 f'(tx)dt).$$

   (c) Prove that the R-modules M and I are isomorphic.

4. Let R be a commutative ring and let $I, J \subset R$ be ideals such that

$$I + J = R \qquad \text{and let} \quad i_1 \in I, \, j_1 \in J \quad \text{satisfy} \quad i_1 + j_1 = 1.$$

   (a) Show that

$$\phi : I \oplus J \longrightarrow R, \qquad (i,j) \mapsto i + j$$

   is a surjective R-module homomorphism with kernel $\mathrm{Ker}(\phi)$ as an R-module isomorphic to $I \cap J = IJ$.

   (b) Show that

$$\psi : I \oplus J \longrightarrow R \oplus IJ, \qquad (i,j) \mapsto (i+j, ij_1 - ji_1)$$

   is an R-module isomorphism.

5. Let R be the ring of polynomial functions on the circle:

$$R = \mathbb{R}[X,Y]/(X^2+Y^2-1)$$

and define

$$x := X + (X^2+Y^2-1),\ y := Y + (X^2+Y^2-1) \in R.$$

Since X^2+Y^2-1 is irreducible (Eisenstein polynomial at $p = Y-1$ in the unique factorization domain $(\mathbb{R}[X])[Y]$), the ideal $(X^2+Y^2-1) \subset \mathbb{R}[X,Y]$ is a prime ideal and hence R is an integral domain. The field of fractions of R we denote by $Q(R)$. In R we define the ideals

$$I := (x-1, y) \qquad \text{and} \qquad J := (x, y-1),$$

it can be verified that I is not a principal ideal (this is an exercise in the *Algebraic Structures* course).

   (a) Prove that in R it holds that

$$(x+y-1)^2 = 2(x-1)(y-1), \quad (x+y-1)(x-y+1) = -2y(y-1).$$

   (b) Define

$$\psi : I \longrightarrow Q(R), \qquad i \mapsto i \cdot \frac{2(y-1)}{x+y-1}.$$

   Verify that ψ is an R-module homomorphism, that ψ is injective, and that $\mathrm{Im}(\psi) = J \subset R \subset Q(R)$.

   (c) Prove that $I + J = R$ and that $IJ = (x+y-1)R \cong R$ (isomorphic as R-modules).

   (d) Conclude (using Exercise 4) that $I \ncong R$, but $I \oplus I \cong R^2$.

6. Take $R = \mathbb{Z}[\sqrt{-5}]$ and $I := (2, 1+\sqrt{-5})$ and $J := (3, 1-\sqrt{-5})$ (so I and J are ideals in R).

   (a) Verify that

$$\psi : I \longrightarrow \mathbb{Q}[\sqrt{-5}], \qquad i \mapsto i \cdot \frac{3}{1+\sqrt{-5}},$$

   is an injective R-module homomorphism with $\mathrm{Im}(\psi) = J \subset R$.

   (b) Show that $I + J = R$ and $IJ = (1-\sqrt{-5})R$.

   (c) Show using Exercise 4 that $I \ncong R$ but $I \oplus I \cong R^2$ (for a different argument compare Example VI.3.5).

7. This exercise provides an example of a ring R such that $R \cong R^2$ as R-modules. Let R be the ring of row-finite matrices with coefficients in the field K. So any $r \in R$ is an infinite matrix $r = (r_{ij})_{i,j \in \mathbb{Z}_{\geq 1}}$, and for every i we have $r_{ij} \neq 0$ for only finitely many j.
   Addition and multiplication are analogous to the usual matrix operations:

$$r + s = t \quad \text{with } t_{ij} := r_{ij} + s_{ij}, \quad \text{and} \quad rs = u \quad \text{with } u_{ij} := \sum_{k=0}^{\infty} r_{ik} s_{kj}$$

   (note that the latter sum is a finite sum since for any i only finitely many r_{ik}'s are nonzero).

   (a) Verify that R is a ring.

   (b) Define $b, c \in R$ by

$$b := (b_{ij}) \quad \text{and} \quad b_{ij} = 1 \text{ for } j = 2(i-1)+1 \text{ and } b_{ij} = 0 \text{ otherwise,}$$

$$c := (c_{ij}), \quad \text{and} \quad c_{ij} = 1 \text{ for } j = 2(i-1)+2 \text{ and } c_{ij} = 0 \text{ otherwise.}$$

   Prove that

$$R \cong Rb \oplus Rc \cong R \oplus R$$

   as R-modules.

8. An R-module $M \neq 0$ is called *simple* if $\{0\}$ and M are the only R-submodules of M. Suppose M is a simple R-module.
   Prove that for every R-module homomorphism $f : M \to M$ either $f = 0$ or f is an R-module isomorphism. Conclude that $\mathrm{End}_R(M)$ is a division ring (so every nonzero element is a unit).
   Determine $\mathrm{End}_R(M)$ in the case $R = M(n,K)$ and $M = K^n$, where K is a field (see Example VI.1.3).
9. Given are two finite dimensional vectorspaces V, W over a field K. Let $\alpha : V \to V$ and $\beta : W \to W$ be linear maps over K, with minimal polynomials m_α and m_β. Define
$$\alpha \oplus \beta : V \oplus W \to V \oplus W \quad \text{by} \quad (v,w) \mapsto \big(\alpha(v), \beta(w)\big).$$
   Prove that if $\gcd(m_\alpha, m_\beta) = 1$ then $m_{\alpha\oplus\beta} = m_\alpha m_\beta$.

VII QUOTIENTS, EXACTNESS, TENSOR PRODUCTS, AND PROJECTIVE MODULES

VII.1 Quotients of modules

Recall that any subgroup N of an abelian group M is a normal subgroup, so the factor group M/N exists. If moreover M is an R-module for some ring R, and $N \subset M$ is an R-submodule, then a natural structure of R-module exists on the factor group M/N. This is explained below.

VII.1.1 Definition. Let R be a ring and let N be a submodule of an R-module M. Then

$$M/N = \{\bar{m} = m + N \subset M : \ m \in M\}$$

obtains the structure of an R-module by defining

$$R \times M/N \longrightarrow M/N, \qquad (r, m+N) \mapsto rm + N.$$

This R-module is simply denoted M/N and it is called the *quotient* of M by N.

VII.1.2 Remark. Of course one needs to check that the map $R \times M/N \to M/N$ used here is well-defined, in other words:

$$\text{if} \quad m_1 + N = m_2 + N \qquad \text{then} \quad rm_1 + N = rm_2 + N.$$

This property indeed holds, since

$$m_1 + N = m_2 + N \implies m_1 - m_2 \in N, \qquad \text{and} \quad rN \subset N$$

(because N is an R-module). Now $m_1 - m_2 \in N$ implies $r(m_1 - m_2) = rm_1 - rm_2 \in N$, which exactly means $rm_1 + N = rm_2 + N$. So $m_1 + N = m_2 + N \Rightarrow rm_1 + N = rm_2 + N$.

One easily verifies that in this way the abelian group obtains the structure of an R-module, as asserted in the definition.

VII.1.3 Examples.

1. Let $R = K$ be a field and let $V \subset K^n$ be a linear subspace of K^n. Then the quotient K^n/V is a vectorspace over K. To make this more explicit, take a basis $f_1, \ldots, f_n$ of K^n by extending a basis $f_1, f_2, \ldots, f_k$ of V. So

$$V = \left\{ \sum_{i=1}^{n} x_i f_i \in K^n : x_i = 0 \text{ for all } i > k \right\}.$$

Define

$$W := \sum_{j=k+1}^{n} K \cdot f_j.$$

Then $K^n = V \oplus W$, which means every $x \in K^n$ can be written in a unique way as $x = v + w$ with $v \in V$, $w \in W$. In particular $x \in V$ if and only if $w = 0$.
So if $x_1 = v_1 + w_1$ and $x_2 = v_2 + w_2$ for some $v_i \in V$ and $w_i \in W$ then

$$x_1 + V = x_2 + V \iff x_1 - x_2 \in V \iff w_1 = w_2.$$

As a consequence

$$K^n/V = \{w + V \subset K^n : w \in W\},$$

and

$$w_1 + V = w_2 + V \Leftrightarrow w_1 = w_2.$$

This means that any residue class $x + V$, by writing $x = v + w$ with $v \in V$ and $w \in W$, can be written in a unique way as $x + V = w + V$ with $w \in W$. In this way the vectorspace K^n/V is identified with the vectorspace W. More precisely, the map

$$f : W \longrightarrow K^n/V, \qquad w \mapsto w + V$$

is an isomorphism of vectorspaces over K. Note that there are many different choices for W, but all of them are K-vectorspaces isomorphic to K^{n-k}.

2. If G is an abelian group and $H \subset G$ a subgroup, then G/H is an abelian group as well and therefore it is a $\mathbb{Z}$-module. The action of $\mathbb{Z}$ we defined in Example VI.1.5 on any abelian group and hence also on G/H, coincides with the action of $\mathbb{Z}$ defined here on the quotient module G/H.
3. If R is a ring and $I \subset R$ is an ideal, then I is a submodule of the R-module R. The R-module structure on the quotient R/I is the same as the one defined in Example VI.1.3 (verify this yourself!).

VII.1.4 Theorem. *Let $f : M \to N$ be an R-module homomorphism. Then:*

(a) *The R-module homomorphism f induces an R-module isomorphism*

$$M/\mathrm{Ker}(f) \xrightarrow{\cong} \mathrm{Im}(f), \qquad m + \mathrm{Ker}(f) \mapsto f(m).$$

(b) *If $K \subset M$ is a submodule, then the* canonical *map*

$$\phi : M \longrightarrow M/K, \qquad m \mapsto m + K$$

is a (surjective) R-module homomorphism with kernel $\mathrm{Ker}(\phi) = K$.

Proof. This is completely analogous to similar homomorphism theorems in Group Theory and in the theory of rings. For this reason the details are left as an exercise for the reader. ∎

VII.2 Hom and exactness

The following terminology for R-module homomorphisms is very common not only in algebra, but also in, for example, topology (in particular homology and homotopy theory) and in differential geometry.

VII.2.1 Definition. A sequence of R-module homomorphisms

$$\ldots \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P \longrightarrow \ldots$$

is called *exact in* N if

$$\mathrm{Im}(f) = \mathrm{Ker}(g).$$

The sequence is called exact if it is exact in all modules appearing in it.

VII.2.2 Remark. A special case of a sequence of R-module homomorphisms is

$$0 \longrightarrow N \xrightarrow{g} P.$$

This sequence is exact in N precisely when g is injective: namely, the image of the map on the left is $\{0\} \subset N$, so exactness (in N) means that $\{0\} = \mathrm{Ker}(g)$, in other words, g is injective.

Similarly, the sequence

$$M \xrightarrow{f} N \longrightarrow 0$$

is exact in N precisely when f is surjective. Indeed, the kernel of the map on the right is all of N, so exactness (in N) means $\mathrm{Im}(f) = N$.

Finally, consider the sequence

$$0 \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P \longrightarrow 0.$$

This sequence being exact means first of all that f is injective, so $f : M \to f(M)$ is an R-module isomorphism. Next, g is surjective and therefore $\mathrm{Ker}(g) = \mathrm{Im}(f) \cong M$, hence g induces an R-module isomorphism $N/\mathrm{Ker}(g) \cong P$. Up to identifications (which means R-module isomorphisms), this means that an exact sequence as above has the form

$$0 \longrightarrow \mathrm{Ker}(g) \hookrightarrow N \overset{g}{\twoheadrightarrow} N/\mathrm{Ker}(g) \longrightarrow 0.$$

If a sequence

$$\ldots \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P \longrightarrow \ldots$$

is exact, then the property $\mathrm{Ker}(g) = \mathrm{Im}(f)$ implies in particular that $g \circ f = 0$. In other words, the composition of any two consecutive maps in an exact sequence, is the zero map. On the other hand, if in a sequence $M \xrightarrow{f} N \xrightarrow{g} P$ we have $g \circ f = 0$ then this implies $\mathrm{Im}(f) \subset \mathrm{Ker}(g)$, which is evidently weaker than the assertion that the sequence is exact at N.

VII.2.3 Definition. Given a ring R, a *diagram* of R-modules is a simple, directed graph in which the vertices represent R-modules M_i, and the (directed) edges represent R-module homomorphisms $M_i \to M_j$.

A diagram of R-modules is called a *commutative diagram* if (for all i, j) every path from M_i to M_j, seen as a composition of R-module homomorphisms, is the same map.

VII.2.4 Examples. A triangle of R-modules and maps

$$\begin{array}{ccc} L & & \\ \downarrow f & \searrow^{g} & \\ M & \xrightarrow{h} & N \end{array}$$

being commutative simply means that $h \circ f = g$.

Similarly, commutativity of a square

$$\begin{array}{ccc} K & \xrightarrow{e} & L \\ \downarrow f & & \downarrow g \\ M & \xrightarrow{h} & N \end{array}$$

means that $h \circ f = g \circ e$.

A famous result featuring commutative diagrams as well as exact sequences is the so-called snake lemma. It appeared in the movie *It's my turn* (1980) where it was proven by actress Jill Clayburgh (1944–2010). We present it here, using the same notations as she did in the movie. To this end we first introduce one more notation.

VII.2.5 Notation. If R is a ring and $f: M \to N$ is an R-module homomorphism, then we write

$$\mathrm{Coker}(f) := N/f(M)$$

for the R-module obtained by taking the quotient of N by the submodule $f(M)$.

VII.2.6 Theorem. *Suppose the diagram of R-modules*

$$\begin{array}{ccccccccc} 0 \to & A & \xrightarrow{f} & B & \xrightarrow{g} & C & \to 0 \\ & \downarrow\alpha & & \downarrow\beta & & \downarrow\gamma & \\ 0 \to & A' & \xrightarrow{f'} & B' & \xrightarrow{g'} & C' & \to 0 \end{array}$$

is commutative, and its two rows

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0 \qquad \textit{and} \qquad 0 \to A' \xrightarrow{f'} B' \xrightarrow{g'} C' \to 0$$

are exact.

Then this induces an exact sequence of R-modules

$$0 \to \mathrm{Ker}(\alpha) \xrightarrow{f} \mathrm{Ker}(\beta) \xrightarrow{g} \mathrm{Ker}(\gamma) \xrightarrow{\delta} \mathrm{Coker}(\alpha) \xrightarrow{\overline{f'}} \mathrm{Coker}(\beta) \xrightarrow{\overline{g'}} \mathrm{Coker}(\gamma) \to 0.$$

Here f and g denote the restrictions to $\mathrm{Ker}(\alpha) \subset A$ and $\mathrm{Ker}(\beta) \subset B$ of $f: A \to B$ resp. $g: B \to C$. Similarly, $\overline{f'}$ and $\overline{g'}$ are the maps induced by $f': A' \to B'$ resp. $g': B' \to C'$. Finally, δ is defined as follows. Given any $c \in \mathrm{Ker}(\gamma) \subset C$, since $g: B \to C$ is surjective, $b \in B$ exists with $g(b) = c$. Then $g'(\beta(b)) = \gamma(g(b)) = \gamma(c) = 0$ by the commutativity of the diagram, hence $\beta(b) \in \mathrm{Ker}(g') = \mathrm{Im}(f')$, so $a' \in A'$ exists with $f'(a') = \beta(b)$. Now put $\delta(c) := a' + \alpha(A) \in \mathrm{Coker}(\alpha)$.

Proof. We first show that the maps in the sequence are well-defined and indeed map to the indicated R-modules.
For (the restriction of) f, if $a \in \mathrm{Ker}(\alpha)$ then $\beta(f(a)) = f'(\alpha(a)) = f'(0) = 0$ where we used that the diagram is commutative. So indeed $f(a) \in \mathrm{Ker}(\beta)$.
The same argument works for the restriction of g.
In the construction of δ, there are two places where potentially a choice is made. One of them is that given $\beta(b) \in \mathrm{Im}(f')$ we choose $a' \in A'$ with $f'(a') = \beta(b)$. However, f' is injective (by the exactness of the lower row in the diagram), so in fact we have a unique a' here. The other choice is the $b \in B$ such that $g(b) = c$. If a different choice $b_1 \in B$ was used, then $g(b) = c = g(b_1)$ hence $b_1 - b \in \mathrm{Ker}(g) = \mathrm{Im}(f)$. This means $a \in A$ exists with $b_1 = b + f(a)$. Then $\beta(b_1) = \beta(b) + \beta(f(a)) = \beta(b) + f'(\alpha(a))$. Hence $\beta(b_1)$ is the image (under f') of the unique element $a' + \alpha(a) \in A'$. In particular the classes of a' and of $a' + \alpha(a)$ in $\mathrm{Coker}(\alpha) = A'/f(A)$ are equal, showing that δ is well-defined.
For the map $\overline{f'}$, first consider the composition $\pi \circ f': A' \xrightarrow{f'} B' \xrightarrow{\pi} B'/\beta(B)$ where π is the canonical map. If $a' \in \alpha(A)$, then write $a' = \alpha(a)$ with $a \in A$. We have, using the commutativity of the diagram that $\pi(f'(\alpha(a))) = \pi(\beta(f(a)) = 0$, since $\beta(f(a)) \in \beta(B)$. Hence $\alpha(A) \subset \mathrm{Ker}(\pi \circ f')$, which implies that $\overline{f'}: \mathrm{Coker}(\alpha) \to \mathrm{Coker}(\beta)$ given by $a' + \alpha(A) \mapsto f'(a) + \beta(B)$ is well-defined.
The same argument shows that $\overline{g'}$ is well-defined.

The fact that all maps in the sequence are R-module homomorphisms, is easily checked from the definitions. So it remains to show that

$$0 \to \mathrm{Ker}(\alpha) \xrightarrow{f} \mathrm{Ker}(\beta) \xrightarrow{g} \mathrm{Ker}(\gamma) \xrightarrow{\delta} \mathrm{Coker}(\alpha) \xrightarrow{\overline{f'}} \mathrm{Coker}(\beta) \xrightarrow{\overline{g'}} \mathrm{Coker}(\gamma) \to 0$$

is exact. This is checked at each of the modules in the sequence, as follows.

Ker(α): since $f : A \to B$ is injective, so is its restriction to Ker(α).

Ker(β): the exactness of $A \xrightarrow{f} B \xrightarrow{g} C$ says $f(A) = \mathrm{Ker}(g)$. So restricting g to Ker(β) one obtains as kernel $\mathrm{Ker}(\beta) \cap f(A)$. Since $g \circ f = 0$ this certainly contains $f(\mathrm{Ker}(\alpha))$. On the other hand, if $b \in \mathrm{Ker}(\beta) \cap f(A)$ then $b = f(a)$ for some $a \in A$ and $\beta(f(a)) = 0$. As $\beta \circ f = f' \circ \alpha$ and f' is injective, this implies $\alpha(a) = 0$. So $a \in \mathrm{Ker}(\alpha)$ and then one concludes $b = f(a) \in f(\mathrm{Ker}(\alpha))$. This shows exactness at Ker(β).

Ker(γ): take $c \in \mathrm{Ker}(\gamma)$ such that $c \in \mathrm{Im}(g)$. This means we have $b \in \mathrm{Ker}(\beta)$ with $g(b) = c$. The fact that $b \in \mathrm{Ker}(\beta)$, says that $\beta(b) = 0$. And $0 \in B'$ is the image of $0 \in A'$ under the map f', so by definition $\delta(c) = 0 + \alpha(A) \in \mathrm{Coker}(\alpha)$, in other words $c \in \mathrm{Ker}(\delta)$. This shows $\mathrm{Im}(g) \subset \mathrm{Ker}(\delta)$. On the other hand, is $\delta(c) = \overline{0}$ for some $c \in \mathrm{Ker}(\gamma)$, then for a $b \in B$ with $g(b) = c$ we have that $\beta(b) = f'(a')$ with $a' \in \alpha(A)$. So $a' = \alpha(a)$ for some $a \in A$, and then $\beta(b) = f'(\alpha(a)) = \beta(f(a))$. As a result $b - f(a) \in \mathrm{Ker}(\beta)$, and we have $g(b - f(a)) = g(b) - g(f(a)) = c - 0 = c$. This shows $\mathrm{Ker}(\delta) \subset \mathrm{Im}(g)$ and finishes the proof of exactness at Ker(γ).

Coker(α): first, we show that $\mathrm{Im}(\delta) \subset \mathrm{Ker}(\overline{f'})$. So take $\overline{a'} \in \mathrm{Coker}(\alpha)$ with $\overline{a'} = \delta(c)$ where $c \in C$ satisfies $\gamma(c) = 0$. By definition $\overline{f'}(\overline{a'}) = f'(a') + \beta(B)$. We have to show that this is zero, in other words, that $f'(a') \in \beta(B)$. To this end, pick $b \in B$ with $g(b) = c$ and $a'' \in A'$ with $f'(a'') = \beta(b)$. By the construction of the map δ, then $\overline{a'} = \delta(c) = \overline{a''}$. Hence $a' = a'' + \alpha(a)$ for some $a \in A$. As a result, $f'(a') = f'(a'') + f'(\alpha(a))$, and this equals $\beta(b) + \beta(f(a)) = \beta(b + f(a))$. So indeed $f'(a') \in \beta(B)$.
Next we show $\mathrm{Ker}(\overline{f'}) \subset \mathrm{Im}(\delta)$. Take $\overline{a'} \in \mathrm{Coker}(\alpha)$ with $\overline{f'}(\overline{a'}) = 0$. By definition this means that $b \in B$ exists with $f'(a') = \beta(b)$. Now put $c := g(b) \in C$. Then $c \in \mathrm{Ker}(\gamma)$ since $\gamma(c) = \gamma(g(b)) = g'(\beta(b)) = g'(f'(a')) = 0$. By the construction of δ we have $\delta(c) = \overline{a'}$. This finished proving exactness at Coker(α).

Coker(β): Given $\overline{b'} \in \mathrm{Coker}(\beta)$, if $\overline{g'}(\overline{b'}) = 0$ then $c \in C$ exists with $g'(b') = \gamma(c)$. Now write $c = g(b)$ for some $b \in B$, then $g'(b' - \beta(b)) = \gamma(c) - \gamma(g(b)) = 0$ hence $a' \in A'$ exists with $f'(a') = b' - \beta(b)$. This implies that $\overline{b'} \in \mathrm{Im}(\overline{f'})$.
Vice versa, is $\overline{b'} \in \mathrm{Im}(\overline{f'})$ then $a' \in A'$ and $a \in A$ exist such that $f'(a') - b' = \alpha(a)$. Hence $g'(b') = g'(f'(a') - g'(\alpha(a)) = 0 - \beta(g(a)) \in \beta(B)$. This means $\overline{b'} \in \mathrm{Ker}(\overline{g'})$.

Coker(γ): here we have to show that $\overline{g'}$ is surjective. So take $\overline{c'} \in \mathrm{Coker}(\gamma)$. Since g' is surjective, $b' \in B'$ exists with $g'(b') = c'$. Then $\overline{g'}(\overline{b'}) = \overline{c'}$.
This finishes the proof of the snake lemma. ∎

VII.2.7 Definition. Given a ring R and (left) R-modules M, N, the set of all R-module homomorphisms $M \to N$ is denoted

$$\mathrm{Hom}(M,N) = \{f : M \to N : \ f \text{ is an } R\text{−module homomorphism}\}.$$

If one wants to emphasize that the homomorphisms are considered over the ring R, then the notation $\mathrm{Hom}_R(M,N)$ is used.

In case $N = M$ one writes $\mathrm{End}_R(M) := \mathrm{Hom}_R(M,M)$, the R-module endomorphisms of M.

The set $\mathrm{Hom}(M,N)$ is in a natural way an abelian group, with

$$(f+g)(m) := f(m)+g(m) \qquad (f,g \in \mathrm{Hom}(M,N),\ m \in M).$$

In the case $M = N$, i.e., considering $\mathrm{End}(M)$, the product given by composing maps: $fg = f \circ g$, provides $\mathrm{End}(M)$ with a ring structure. In fact this is a subring of the ring consisting of all group homomorphisms $M \to M$.

In case R is *commutative*, the abelian group $\mathrm{Hom}(M,N)$ obtains the structure of an R-module by defining

$$(rf)(m) := rf(m), \qquad (f \in \mathrm{Hom}(M,N),\ m \in M,\ r \in R).$$

Commutativity of R is essential, since we want $rf \in \mathrm{Hom}(M,N)$. This means that rf should be R-linear. Now

$$(rf)(sm) := r \cdot (f(sm)) = (rs) \cdot f(m),$$

and this equals $(sr) \cdot f(m)$ in case R is *commutative*, but not necessarily otherwise. One readily checks that indeed in this way $\mathrm{Hom}(M,N)$ is an R-module in case R is commutative.

In de remainder of this section we will assume the ring R to be *commutative*.

(Most of the results presented here also hold for non-commutative rings, provided $\mathrm{Hom}(M,N)$ is considered as an abelian group only.)

VII.2.8 Examples. 1. Let K be a field and take $M = N = K^n$. Then

$$\mathrm{End}_K(K^n) := \mathrm{Hom}_K(K^n,K^n) \cong M(n,K),$$

with $M(n,K)$ the K-module consisting of $n \times n$ matrices with coefficients in K, and the usual addition and scalar multiplication of matrices. Namely, any $\alpha \in \mathrm{Hom}_K(K^n,K^n)$ is by definition a K-linear map.

2. Let R be a ring and M an R-module. Then

$$\mathrm{ev}_1 : \mathrm{Hom}_R(R,M) \xrightarrow{\cong} M, \qquad f \mapsto f(1),$$

defines an isomorphism of R-modules. Injectivity of ev_1 follows from $f(r) = rf(1)$. Namely, is $\mathrm{ev}_1(f) := f(1) = 0$ then $f(r) = rf(1) = r \cdot 0 = 0$ for all $r \in R$, showing that $f = 0$. To show that ev_1 is surjective, note that for any $m \in M$ the map

$$f_m : R \longrightarrow M \qquad f_m(r) := rm$$

defines an R-module homomorphism. It satisfies $\mathrm{ev}_1(f_m) = m$. It remains to show that ev_1 is an R-module homomorphism. This is immediate:

$$\begin{array}{rcll} \mathrm{ev}_1(f+g) & := & (f+g)(1) & := f(1)+g(1) = \mathrm{ev}_1(f)+\mathrm{ev}_1(g), \\ \mathrm{ev}_1(rf) & := & rf(1) & = r\mathrm{ev}_1(f). \end{array}$$

So indeed ev_1 defines an isomorphism of modules.

3. For every $n \in \mathbb{Z}_{\geq 1}$ it holds that

$$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z},\mathbb{Z}) = 0.$$

Indeed, if $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}$ is $\mathbb{Z}$-linear, then

$$0 = f(\overline{0}) = f(\overline{n}) = f(n \cdot \overline{1}) = nf(\overline{1}),$$

and since $\mathbb{Z}$ is an integral domain and $n \neq 0$, this implies $f(\overline{1}) = 0$. Hence $f(\overline{a}) = af(\overline{1}) = a \cdot 0 = 0$ for all $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ which shows that $f = 0$. In words: every element in $\mathbb{Z}/n\mathbb{Z}$ has finite order and its image under f therefore has finite order as well. As a consequence, this image is 0 since 0 is the only element of finite order in $\mathbb{Z}$.

We will now consider $\mathrm{Hom}_R(A,-)$; in other words we fix the R-module A and consider the recipe that on input any R-module M, outputs the R-module $\mathrm{Hom}_R(A,M)$. As before, all rings R considered here will be assumed commutative.

VII.2.9 Definition. If A is an R-module, and $f \in \mathrm{Hom}(M,N)$, then composing any R-module homomorphism $\phi\colon A \to M$ with $f : M \to N$ results in an R-module homomorphism denoted $f_*(\phi) := f \circ \phi\colon A \to N$:

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ & {\scriptstyle\phi}\nwarrow\ \nearrow{\scriptstyle f_*(\phi)} & \\ & A & \end{array} \qquad \text{with} \quad f_* : \mathrm{Hom}(A,M) \longrightarrow \mathrm{Hom}(A,N), \qquad \phi \mapsto f \circ \phi.$$

One also uses the somewhat suggestive notation $\mathrm{Hom}(A,f)$ for f_*.

Using the definitions it is not hard to check that indeed f_* is an R-module homomorphism, which in this case means that

$$f_*(\phi+\psi) = f_*(\phi) + f_*(\psi), \qquad \text{and} \qquad f_*(r\phi) = r f_*(\phi)$$

for all ϕ, $\psi \in \mathrm{Hom}(A,M)$ and all $r \in R$.

We will now consider the following problem. Suppose $f : M \to N$ is surjective. Does it follow that $f_* : \mathrm{Hom}(A,M) \to \mathrm{Hom}(A,N)$ is surjective as well? R-modules A for which the answer is affirmative, so with the property

$$f : M \longrightarrow N \quad \text{surjective} \quad \Longrightarrow \quad f_* : \mathrm{Hom}(A,M) \longrightarrow \mathrm{Hom}(A,N) \quad \text{surjective}$$

will be studied in more detail in Section VII.4.

(We will see in Theorem VII.2.12 that if one replaces 'surjective' by 'injective'in the question above, then for all R-modules A the answer is affirmative. Note also that if f is not surjective, then Exercise 1 on page 74 shows that already for $A = R$ the map f_* is also not surjective.

VII.2.10 Example. This example shows that f_* is not necessarily surjective even if f is surjective. Consider $n \in \mathbb{Z}_{\geq 2}$ and

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\pi} & \mathbb{Z}/n\mathbb{Z} \\ & {\scriptstyle\tilde{h}}\nwarrow\ \nearrow{\scriptstyle \mathrm{id}_{\mathbb{Z}/n\mathbb{Z}}} & \\ & \mathbb{Z}/n\mathbb{Z} & \end{array} \qquad \text{with} \quad f : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \quad \text{given by} \quad a \mapsto a + n\mathbb{Z}.$$

If π_* were surjective, then $\tilde{h} : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}$ exists with $\pi \circ \tilde{h} = \mathrm{id}_{\mathbb{Z}/n\mathbb{Z}}$.

By Example VII.2.8(3) we know $\mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) = 0$, hence $\tilde{h} = 0$. But then $\mathrm{id}_{\mathbb{Z}/n\mathbb{Z}} = \pi \circ \tilde{h} = 0$, a contradiction. So π_* is not surjective. ∎

VII.2.11 Example. We now show: if $A = R$ and $f : M \to N$ is a surjective R-module homomorphism, then $f_* : \mathrm{Hom}(R,M) \to \mathrm{Hom}(R,N)$ is surjective as well.

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle\tilde{h}}\ \nwarrow & \nearrow & {\scriptstyle h} \\ & R & \end{array}$$

To show this, take any $h \in \mathrm{Hom}(R,N)$. We want to find $\tilde{h} \in \mathrm{Hom}(R,M)$ such that $f_*(h) = f \circ \tilde{h} = h$. Since f is surjective, $m \in M$ exists with $f(m) = h(1)$. Now define

$$\tilde{h} : R \to M \qquad \text{by} \qquad \tilde{h}(r) := rm.$$

Then $\tilde{h} \in \mathrm{Hom}(R,M)$ (see Example VII.2.8(2)) and moreover

$$(f_*\tilde{h})(r) := f\tilde{h}(r) = f(rm) = rf(m) = rh(1) = h(r)$$

for all $r \in R$. So indeed $f_*\tilde{h} = h$, showing that f_* is surjective in this case. ∎

VII.2.12 Theorem. *Let R be a commutative ring and suppose*

$$0 \longrightarrow M \xrightarrow{f} N \xrightarrow{g} P$$

is an exact sequence of R-modules.
For any R-module A it holds that

$$0 \longrightarrow \mathrm{Hom}(A,M) \xrightarrow{f_*} \mathrm{Hom}(A,N) \xrightarrow{g_*} \mathrm{Hom}(A,P)$$

is an exact sequence of R-modules.
In particular: if $f: M \to N$ is injective, then so is $f_*: \mathrm{Hom}(A,M) \to \mathrm{Hom}(A,N)$.

VII.2.13 Remark. One usually expresses Theorem VII.2.12 by asserting that the map $\mathrm{Hom}(A,-)$ is *left-exact*.

Proof. (of Theorem VII.2.12.) Given are an exact sequence as above and an R-module A.

Exactness at $\mathrm{Hom}(A,M)$: suppose $\phi \in \mathrm{Hom}(A,M)$ satisfies $\phi \neq 0$. Then $a \in A$ exists with $\phi(a) \neq 0$. As $\phi(a) \in M$ and $f: M \to N$ is injective, $(f_*\phi)(a) := f(\phi(a)) \neq 0$ so $f_*\phi \neq 0$. Hence f_* is injective.

Exactness at $\mathrm{Hom}(A,N)$: we must show $\mathrm{Im}(f_*) = \mathrm{Ker}(g_*)$.
'$\subset$': take an arbitrary $\psi \in \mathrm{Im}(f_*)$. We may write $\psi = f_*(\phi)$ for some $\phi \in \mathrm{Hom}(A,M)$. For any $a \in A$ now $\psi(a) = f(\phi(a)) \in \mathrm{Im}(f)$. By assumption $\mathrm{Im}(f) = \mathrm{Ker}(g)$, hence $g\psi(a) = 0$ for all $a \in A$. This shows $g_*(\psi) = 0$, proving that $\mathrm{Im}(f_*) \subset \mathrm{Ker}(g_*)$.
'$\supset$': let $\psi \in \mathrm{Hom}(A,N)$ satisfy $g_*(\psi) = 0$. We will construct $\phi \in \mathrm{Hom}(A,M)$ such that $f_*(\phi) = \psi$. The condition $g_*(\psi) = 0$ means that $g\psi(a) = 0$ for all $a \in A$, i.e., $\psi(a) \in \mathrm{Ker}(g)$ for all $a \in A$. Moreover by assumption $\mathrm{Ker}(g) = \mathrm{Im}(f)$. We also have that f is injective, so $f: M \xrightarrow{\cong} \mathrm{Im}(f)$. As a result, an R-module isomorphism

$$M \xleftarrow{h} \mathrm{Im}(f) \quad (\subset N), \qquad \text{with} \quad f \circ h = id_{\mathrm{Im}(f)}$$

exists. Since $im(\psi) \subset im(f)$, the composition $\phi := h \circ \psi: A \to M$ is a well defined R-module homomorphism. One finds for all $a \in A$ that

$$(f_*\phi)(a) := f(\phi(a)) = f(h(\psi(a))) = \psi(a), \quad \text{hence} \quad f_*(\phi) = \psi$$

(here we used that $\psi(a) \in \mathrm{Ker}(g) = \mathrm{Im}(f)$ and $f \circ h = id_{\mathrm{Im}(h)}$). So $\mathrm{Ker}(g_*) \subset \mathrm{Im}(f_*)$.
This concludes the proof of the theorem. ∎

VII.3 Tensor products

In linear algebra, and in various applications of linear algebra such as coding theory, inner products and generalizations of inner products are an important notion. Tensor products, which will be introduced in this section, provide a general framework for this. They are in particular used in representation theory and in differential geometry, and in applications of these areas in theoretical physics.

VII.3.1 Definition. If R is a ring and M,N,T are R-modules, then a map

$$b: M \times N \longrightarrow T$$

is called *bilinear* (or R-bilinear) if for every $m \in M$ and every $n \in N$ the maps $M \to T$ and $N \to T$ given by $x \mapsto b(x,n)$ resp. $y \mapsto b(m,y)$ are R-module homomorphisms.

VII.3.2 Example. Take $n \in \mathbb{Z}_{\geq 1}$. Let R be a commutative ring and let $M = N = R^n$ be the free R-module of rank n. Then

$$b : R^n \times R^n \to R \qquad \text{given by} \quad b((x_1,\ldots,x_n),(y_1,\ldots,y_n)) = \sum x_j y_j$$

is bilinear. ∎

VII.3.3 Definition. If R is a ring and M, N are R-modules, then a *tensor product* of M and N is a pair (T, β) in which T is an R-module and $\beta : M \times N \to T$ is a bilinear map, such that the following holds:
given any R-bilinear map $b : M \times N \to S$ for some R-module S, there exists a unique R-module homomorphism $f : T \to S$ such that $b = f \circ \beta$.

One can visualize this definition by means of diagrams:

$$\begin{array}{ccc} M \times N & \stackrel{b}{\to} & S \\ \downarrow \beta & & \\ T & & \end{array} \qquad \Rightarrow \qquad \begin{array}{ccc} M \times N & \stackrel{b}{\to} & S \\ \downarrow \beta & \nearrow \exists! f & \\ T & & \end{array}$$

The first thing we will see concerning a tensor product, is that if it exists then it is unique (in fact up to a unique(!) isomorphism). This should be compared with Theorem VI.3.8, where a similar argument for direct sums rather than tensor products is given.

VII.3.4 Theorem. *Suppose M, N are R-modules and (T_1, β_1) and (T_2, β_2) are tensor products of M and N. Then there is a unique R-module isomorphism $f : T_1 \to T_2$ such that $\beta_2 = f \circ \beta_1$.*

Proof. Using that (T_1, β_1) is a tensor product of M and N, the diagram

$$\begin{array}{ccc} M \times N & \stackrel{\beta_2}{\to} & T_2 \\ \downarrow \beta_1 & & \\ T_1 & & \end{array}$$

yields a unique $f_1 : T_1 \to T_2$ with $\beta_2 = f_1 \circ \beta_1$. Interchanging the roles of T_1 and T_2 one obtains $f_2 : T_2 \to T_1$ with $\beta_1 = f_2 \circ \beta_2$. As a result, $\beta_2 = (f_1 \circ f_2) \circ \beta_2$. However, starting from

$$\begin{array}{ccc} M \times N & \stackrel{\beta_2}{\to} & T_2 \\ \downarrow \beta_2 & & \\ T_2 & & \end{array}$$

and the assumption that (T_2, β_2) is a tensor product, the unique arrow $T_2 \to T_2$ making the above diagram commutative is the identity map. So $f_1 \circ f_2 = \mathrm{id}_{T_2}$. Again interchanging the roles of T_1 and T_2, one finds in the same way that $f_2 \circ f_1 = \mathrm{id}_{T_1}$. So f_1 and f_2 are isomorphisms as desired.

To see that f_1 is unique, suppose that also $f_1' : T_1 \to T_2$ is an isomorphism with $\beta_2 = f_1' \circ \beta_1$. Then $\beta_1 = f_1'^{-1} \circ \beta_2$ and hence $\beta_2 = (f_1 \circ f_1'^{-1}) \circ \beta_2$. As above, since (T_2, β_2) is a tensor product this implies $f_1 \circ f_1'^{-1} = \mathrm{id}_{T_2}$ and hence $f_1 = f_1'$. ∎

VII.3.5 Notation. The (if it exists at all) by Theorem VII.3.4 *unique* pair (T, β) which is a tensor product of the R-modules M and N, we will from now on denote by $M \otimes_R N$ (or simply $M \otimes N$ if it is clear from the context which ring R is considered).

The bilinear map $\beta : M \times N$ will be written as $(m, n) \mapsto m \otimes n$.

It remains to show existence of a tensor product. The next result claims exactly that.

VII.3.6 Theorem. *For any R-modules M and N a tensor product (T,β) exists.*

Proof. One constructs an abelian group $(T,+,0)$ as follows. Start with the free abelian group F with (independent) generators denoted $m\times n$ for $m\in M$, $n\in N$. So

$$F\cong \bigoplus_{(m,n)\in M\times N}\mathbb{Z}.$$

Next, take the subgroup $S\subset F$ generated by all elements

$$(m\times n)+(m'\times n)-((m+m')\times n),\quad (m\times n)+(m\times n')-(m\times(n+n')),$$
$$((rm)\times n)-(m\times(rn)),\quad (0\times n)$$

with $m,m',0\in M$ and $n,n'\in N$ and $r\in R$. Let T be the factor group: $T:=F/S$. The class $(m\times n)$ mod $S\in T$ is denoted $m\otimes n$.

One obtains a scalar multiplication by elements of R on F and on S (and therefore on T) by defining

$$r\cdot\left(\sum_j m_j\times n_j\right):=\sum_j(r\cdot m_j)\times n_j.$$

This makes T into an R-module. The map

$$\beta\colon M\times N\to T\qquad \text{given by}\quad \beta(m,n):=m\otimes n=(m\times n)\bmod S$$

is by the definition of S bilinear. The fact that indeed (T,β) defines a tensor product is left as an exercise to the reader. ∎

The *construction* of a tensor product as given in the proof of Theorem VII.3.6 is rarely needed in showing properties of the tensor product. Instead, one shows that a given module is the tensor product of M and N by showing that the module satisfies Definition VII.3.3. Then the unicity of the tensor product (Theorem VII.3.4) shows that indeed the given module is the tensor product. We illustrate this by some simple results and examples.

VII.3.7 Proposition. *Let R be a ring and let M and N be R-modules. Then*

$$M\otimes_R N\cong N\otimes_R M.$$

Proof. It suffices to show that $M\otimes_R N$ has the defining property of a tensor product of N and M.

Firstly, $\beta\colon N\times M\to M\otimes_R N$ given by $\beta(n,m)=m\otimes n$ is bilinear. Now suppose $b\colon N\times M\to T$ is any bilinear map. Then $\tilde{b}\colon M\times N\to T$ defined as $\tilde{b}(m,n)=b(n,m)$ is bilinear as well, so because $M\otimes_R N$ is a tensor product of M and N, a unique R-homomorphism $f\colon M\otimes_R N\to T$ exists with $\tilde{b}=f\circ\beta$. Then also $b=f\circ\beta$, and the uniqueness of an f with the latter property is obvious. ∎

VII.3.8 Proposition. *Let R be a ring and let $f\colon M_1\to M_2$ and $g\colon N_1\to N_2$ be two R-module homomorphisms. Then a unique R-module homomorphism*

$$M_1\otimes_R N_1\to M_2\otimes_R N_2$$

exists with $m_1\otimes n_1\mapsto f(m_1)\otimes g(n_1)$ for all $m_1\in M_1$ and $n_1\in N_1$.

Proof. Define $b\colon M_1\times N_1\to M_2\otimes_R N_2$ by $b(m_1,n_1)=f(m_1)\otimes g(n_1)$. Then b is bilinear, hence by the definition of tensor product a unique R-module homomorphism as desired exists. ∎

VII.3.9 Proposition. *If R is a unitary ring and M is an R-module, then $R \otimes_R M \cong M$.*

Proof. Define $\beta\colon R \times M \to M$ by $\beta(r,m) = rm$. Clearly β is bilinear. We claim that (M,β) satisfies the definition of a tensor product for R and M. Indeed, take any R-module N and a bilinear map $b\colon R \times M \to N$. Define $f\colon M \to N$ by $f(m) = b(1,m)$. Then for any $m \in M$ and any $r \in R$ one has $f(\beta(r,m)) = f(rm) = b(1,rm) = b(r,m)$ since b is bilinear. So $f \circ \beta = b$. If also $f' \circ \beta = b$, then taking any $m \in M$ we have $f(m) = b(1,m) = f'(\beta(1,m)) = f'(m)$, so $f = f'$. Hence f is unique, proving the claim. ∎

VII.3.10 Remark. The proof of Proposition VII.3.9 in fact shows that $r \otimes m \to rm$ defines an isomorphism of R-modules $R \otimes_R M \to M$. This is used in the proof of the next result.

VII.3.11 Proposition. *Let K be a field and let V, W be vectorspaces over K with bases $\{e_i : i \in I\}$ and $\{f_j : j \in J\}$, respectively. Then the tensor product $V \otimes_K W$ is a K-vectorspace with basis $\{e_i \otimes f_j : i \in I, j \in J\}$.*

Proof. By assumption $V \cong \oplus_{i\in I} K$ and $W \cong \oplus_{j\in J} K$. Hence

$$V \otimes_K W \cong \left(\oplus_{i\in I}\ K\right) \otimes_K \left(\oplus_{j\in J} K\right).$$

Now define

$$b\colon \left(\oplus_{i\in I}\ K\right) \times \left(\oplus_{j\in J} K\right) \longrightarrow \oplus_{(i,j)\in I\times J} K \otimes_K K \cong \oplus_{(i,j)\in I\times J} K$$

by $b((x_i)_{i\in I},(y_j)_{j\in J}) = (x_i \otimes y_j)_{(i,j)\in I\times J}$. Since b is bilinear, a (unique)

$$f\colon \left(\oplus_{i\in I}\ K\right) \otimes_K \left(\oplus_{j\in J} K\right) \to \oplus_{(i,j)\in I\times J} K \otimes_K K$$

exists with $b((x_i)_{i\in I},(y_j)_{j\in J}) = f((x_i)_{i\in I} \otimes (y_j)_{j\in J})$. In fact f defines an isomorphism. Since as a special case of Proposition VII.3.9 we have $K \otimes_K K \cong K$, the result follows. ∎

VII.3.12 Remark. In the special case of Proposition VII.3.11 that $\dim_K(V) = n < \infty$ and $\dim_K(W) = m < \infty$, the tensor product $V \otimes_K W$ is a vectorspace of dimension $n \cdot m$.

If in this case $f\colon V \to V$ and $g\colon W \to W$ are K-linear maps, then Proposition VII.3.8 combines these maps into a K-linear map $V \otimes_K W \to V \otimes_K W$. The latter map is usually denoted $f \otimes g$. If f is given by an $n \times n$ matrix and g by an $m \times m$ matrix, then $f \otimes g$ is given by an $nm \times nm$ matrix. With respect to the bases as described in Proposition VII.3.11 the latter matrix is called the *Kronecker product* of the matrices for f and g.

The final result we will present here concerning tensor products, shows a relation between $(T \otimes_R -)$, so the map assigning to an R-module N the R-module $T \otimes_R N$, and $\mathrm{Hom}_R(T,-)$. Recall that for the latter to be a map from R-modules to R-modules, we require that R is commutative.

The result we are about to describe is usually referred to as the statement that $(T \otimes_R -)$ and $\mathrm{Hom}_R(T,-)$ are *adjoint*. To understand this terminology, suppose that V is a vector space over $\mathbb{R}$, equipped with an inner product $\langle\cdot,\cdot\rangle$. Then two linear maps $\varphi, \varphi^*\colon V \to V$ are called adjoint if $\langle\varphi(v),w\rangle = \langle v,\varphi^*(w)\rangle$ for all $v,w \in V$. Now replace "vectors in V" by "R-modules", and "taking the inner product" by "taking $\mathrm{Hom}_R(-,-)$", and the maps φ, φ^* by $(T \otimes_R -)$ respectively $\mathrm{Hom}_R(T,-)$. In this way the following result obtained the name *adjointness of Hom and tensor*.

VII.3.13 Theorem. *Let R be a commutative ring and let T be an R-module. Then for any pair M,N of R-modules one has*

$$\mathrm{Hom}_R(T\otimes_R M,N)\cong \mathrm{Hom}_R(M,\mathrm{Hom}_R(T,N))$$

as R-modules.

Proof. Write $\mathrm{Bilin}_R(T\times M,N)$ for the set of all bilinear maps $T\times M\to N$. In fact this is an R-module with respect to pointwise addition and multiplication by elements of R, so

$$(b+b')(t,m):=b(t,m)+b'(t,m)\qquad\text{and}\quad (rb)(t,m)=r\cdot b(t,m).$$

The definition of a tensor product implies that every element of $\mathrm{Bilin}_R(T\times M,N)$ corresponds to an element of $\mathrm{Hom}_R(T\otimes_R M,N)$, and in this way

$$\mathrm{Bilin}_R(T\times M,N)\cong \mathrm{Hom}_R(T\otimes_R M,N)$$

as R-modules. Hence it suffices to show that

$$\mathrm{Bilin}_R(T\times M,N)\cong \mathrm{Hom}_R(M,\mathrm{Hom}_R(T,N)),$$

which is done as follows. Given $b\in \mathrm{Bilin}_R(T\times M,N)$, which means a bilinear map $b\colon T\times M\to N$, the map $b(-,m)\colon T\to N$ is in $\mathrm{Hom}_R(T,N)$. Vice versa, is $f\in\mathrm{Hom}_R(M,\mathrm{Hom}_R(T,N))$ then $(t,m)\mapsto f(m)(t)$ defines a bilinear map $T\times M\to N$.

A straightforward verification shows that

$$b\mapsto [m\mapsto b(-,m)]$$

and

$$f\mapsto [(t,m)\mapsto f(m)(t)]$$

are each other's inverse, and moreover these maps are R-module homomorphisms. This proves the theorem. ∎

VII.4 Projective modules

In Example VII.2.10 we saw that $\mathrm{Hom}(A,-)$ is in general not (right-)exact, in other words an exact sequence $M\to N\to 0$ does not necessarily give rise to an exact sequence $\mathrm{Hom}(A,M)\to\mathrm{Hom}(A,N)\to 0$. And in Section VI.3 we saw examples of modules M,N that are not free, but $M\oplus N$ is a free module (the examples VI.3.5 and VI.3.6 even have $M=N$).

In this section the notion 'projective module' (Definition VII.4.1) is introduced, and it is shown that $\mathrm{Hom}(P,-)$ is exact precisely when P is projective (see Theorem VII.4.5). Moreover it will be shown that P being projective if and only if a module Q exists such that $P\oplus Q$ is a free module (see Theorem VII.4.6).

VII.4.1 Definition. An R module P is called *projective* if for every surjective R-module homomorphism $M\xrightarrow{f}N$ and for every R-module homomorphism $h\colon P\to N$, so

$$\begin{array}{ccccc} M & \xrightarrow{f} & N & \to 0, \\ & \nearrow_{h} & & \\ P & & & \end{array}$$

there exists an R-module homomorphism $\tilde{h}\colon P\to M$, so

$$\begin{array}{ccccc} M & \xrightarrow{f} & N & \to 0, & \\ {\scriptstyle\tilde{h}}\uparrow & \nearrow_{h} & & & \text{such that}\quad f\tilde{h}=h. \\ P & & & & \end{array}$$

The next three lemmas, especially VII.4.2 and VII.4.4, assist in appreciating this definition.

VII.4.2 Lemma. *Any free R-module F is projective.*

Proof. By definition $F \cong \oplus_{i\in I} R$ for some non-empty set I. Now given and exact sequence $M \xrightarrow{f} N \to 0$ and $h : F \to N$, one constructs $\tilde{h}$ as follows. Put

$$n_i := h(e_i), \qquad \text{and choose} \quad m_i \in M \quad \text{with} \quad f(m_i) = n_i,$$

such m_i exist since f is surjective. Since h is an R-module homomorphism, we have $h(\sum x_i e_i) = \sum x_i h(e_i) = \sum x_i n_i$. Next, define

$$\tilde{h} : F \longrightarrow M, \qquad \text{by} \quad \sum_{i\in I} x_i e_i \mapsto \sum_{i\in I} x_i m_i.$$

Then $\tilde{h}$ is an R-module homomorphism and $f\tilde{h} = h$. This proves the lemma. ∎

VII.4.3 Lemma. *Let R be a ring and let M, N be R-modules. Suppose given two R-module homomorphisms*

$$M \xrightarrow{f} N, \qquad \text{and} \quad M \xleftarrow{g} N \qquad \text{such that} \quad f \circ g = \mathrm{id}_N.$$

Then f is surjective (one says that g splits *the exact sequence $M \xrightarrow{f} N \to 0$). We have*

$$M \cong \mathrm{Im}(g) \oplus \mathrm{Ker}(f), \qquad \text{and} \quad \mathrm{Im}(g) \cong N.$$

Proof. We claim that $\mathrm{Im}(g) \cap \mathrm{Ker}(f) = \{0\}$ and that $\mathrm{Im}(g) + \mathrm{Ker}(f) = M$. If this holds, then the map

$$\mathrm{Im}(g) \oplus \mathrm{Ker}(f) \longrightarrow M, \qquad (a, b) \mapsto a + b$$

is an isomorphism of R-modules (as the reader should verify!).

We first show $\mathrm{Im}(g) \cap \mathrm{Ker}(f) = \{0\}$. Take $x \in \mathrm{Im}(g) \cap \mathrm{Ker}(f)$, then

$$\left.\begin{array}{l} \exists n \in N : \ g(n) = x \\ \qquad\qquad\ \ f(x) = 0 \end{array}\right\} \Longrightarrow n = f(g(n)) = f(x) = 0,$$

and therefore $x = g(n) = g(0) = 0$.

Now take $m \in M$, then

$$gf(m) \in \mathrm{Im}(g) \qquad \text{and} \qquad m = gf(m) + (m - gf(m)),$$

moreover it holds that

$$f(m - gf(m)) = f(m) - fgf(m) = f(m) - f(m) = 0,$$

and therefore $m - gf(m) \in \mathrm{Ker}(f)$. One concludes that every $m \in M$ can be written as a sum of the element $gf(m) \in \mathrm{Im}(g)$ and the element $m - gf(m) \in \mathrm{Ker}(f)$. This shows the lemma. ∎

VII.4.4 Lemma. *If P is a projective R-module and $M \xrightarrow{f} P \longrightarrow 0$ is a surjective R-module homomorphism, then $M \cong P \oplus \mathrm{Ker}(f)$.*

Proof. Use the definition of projective with $N = P$ and $h = \mathrm{id}_P : P \to P = N$ to obtain $\tilde{h} : P \to M$

$$\begin{array}{ccccc} M & \xrightarrow{f} & P & \to 0 & \\ {\scriptstyle \tilde{h}}\uparrow & \nearrow_{\cong} & & & \text{with} \quad f\tilde{h} = h = \mathrm{id}_P. \\ P & & & & \end{array}$$

In particular one concludes that $\tilde{h}$ splits the exact sequence $N \xrightarrow{f} P \to 0$. The lemma now follows from Lemma VII.4.3. ∎

VII.4.5 Theorem. *If P is a projective R-module, then* $\mathrm{Hom}(P,-)$ *is right exact. This means the following. If*

$$A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is an exact sequence of R-modules, then

$$\mathrm{Hom}(P,A) \xrightarrow{f_*} \mathrm{Hom}(P,B) \xrightarrow{g_*} \mathrm{Hom}(P,C) \longrightarrow 0$$

is also an exact sequence of R modules.
Moreover, if an R-module A is not projective, then $\mathrm{Hom}(A,-)$ *is not right exact.*
In conclusion, an R-module P is projective precisely when $\mathrm{Hom}(P,-)$ *is right exact.*

Proof. Given are an exact sequence and a projective module P as in the statement of the theorem.

Exactness at $\mathrm{Hom}(P,C)$: this means we have to verify that g_* is surjective. This follows immediately from the definition of projective: indeed, take $h \in \mathrm{Hom}(P,C)$. Then $\tilde{h} \in \mathrm{Hom}(P,C)$ exists with $f\tilde{h} = h$. This says $f_*(\tilde{h}) = h$, proving surjectivity.

Exactness at $\mathrm{Hom}(P,B)$: we have to show $\mathrm{Im}(f_*) = \mathrm{Ker}(g_*)$.
'$\subset$': let $\psi \in \mathrm{Im}(f_*)$, then one may write $\psi = f_*(\phi)$ for some $\phi \in \mathrm{Hom}(P,A)$. By definition, for any $x \in P$ one has $\psi(x) = f(\phi(x)) \in \mathrm{Im}(f)$. By assumption $\mathrm{Im}(f) = \mathrm{Ker}(g)$, hence $g(\psi(x)) = 0$. This implies $g_*(\psi) = 0$, and therefore $\mathrm{Im}(f_*) \subset \mathrm{Ker}(g_*)$.
'$\supset$': let $\psi \in \mathrm{Hom}(P,B)$ satisfy $g_*(\psi) = 0$. We will construct $\phi \in \mathrm{Hom}(P,A)$ with $f_*(\phi) = \psi$. The condition $g_*\psi = 0$ means

$$\mathrm{Im}(\psi) \subset \mathrm{Ker}(g) = \mathrm{Im}(f).$$

Furthermore $A \xrightarrow{f} \mathrm{Im}(f) \longrightarrow 0$ is an exact sequence. Considering ψ as an R-module homomorphism $\psi\colon P \to \mathrm{Im}(f)$, the definition of projective (with $h := \psi$, and $M := A$, $N := \mathrm{Im}(f)$) implies that $\phi\colon P \to A$ exists with $f\phi = \psi$. This means $f_*(\phi) = \psi$. As a consequence, $\mathrm{Ker}(g_*) \subset \mathrm{Im}(f_*)$.

Finally we prove the remaining assertion of the theorem. If A is not projective, then an exact sequence $M \xrightarrow{f} N \to 0$ and an $h \in \mathrm{Hom}(A,M)$ exist, for which there is no $\tilde{h} \in \mathrm{Hom}(A,N)$ such that $f\tilde{h} = h$. This means that $\mathrm{Hom}(A,-)$, applied to the exact sequence

$$\mathrm{Ker}(f) \hookrightarrow M \xrightarrow{f} N \longrightarrow 0,$$

yields a sequence in which $f_*\colon \mathrm{Hom}(A,M) \to \mathrm{Hom}(A,N)$ is *not* surjective. Hence $\mathrm{Hom}(A,-)$ is not right exact. This finishes the proof. ∎

VII.4.6 Theorem. *Let R be a ring and let P be an R-module. Then:*
P is projective if and only if
an R-module Q exists with $P \oplus Q = F$ for some free R-module F.

Proof. '$\Leftarrow$': Suppose $P \oplus Q = F$ is a free R-module. Given a diagram

$$\begin{array}{ccccc} M & \xrightarrow{f} & N & \to & 0, \\ & \nearrow_{h} & & & \\ P & & & & \end{array}$$

in which f is surjective, we must construct $\tilde{h}$ such that $f\tilde{h} = h$. Define

$$g\colon F = P \oplus Q \longrightarrow N, \qquad g(p,q) := h(p) \quad (p \in P,\, q \in Q).$$

Then $g \restriction P = h$ (the restriction of g to $P \cong P \oplus (0) \subset F$). By Lemma VII.4.2 F is projective, hence $\tilde{g}$ exists with

$$\begin{array}{ccccc} & & M & \xrightarrow{f} & N \quad \to 0 \\ & & {\scriptstyle \tilde{g}}\uparrow & \nearrow_{g} & \\ P \oplus Q = F & & & & \end{array} \qquad \text{such that} \quad f\tilde{g} = g.$$

Now define $\tilde{h} := \tilde{g} \restriction P$. Then $f\tilde{h} = h$ follows from $f\tilde{g} = g$ by restricting to P.
'$\Rightarrow$': Let P be a projective R-module. We construct a free R-module F by using P as index set:

$$F := \bigoplus_{p \in P} R.$$

Now define

$$f : F \to P, \qquad (x_p)_{p \in P} \mapsto \sum_{p \in P} x_p \cdot p.$$

Clearly f is an R-module homomorphism. Moreover f is surjective since $f((x_q)_{q \in P}) = p$ if one takes $x_q = 0$ for all $q \neq p$ and $x_p = 1$. Lemma VII.4.4 then implies $F = P \oplus \mathrm{Ker}(f)$, hence one can take $Q = \mathrm{Ker}(f)$. This proves Theorem VII.4.6. ∎

VII.5 Exercises

1. Let R be a commutative (unitary) ring and suppose $f : M \to N$ is a *non*-surjective R-module homomorphism.
   Show that $f_* : \mathrm{Hom}(R,M) \longrightarrow \mathrm{Hom}(R,N)$ is not surjective.
2. Let R be a commutative ring and consider R-modules M, N, P, Q. Construct R-module isomorphisms

$$\mathrm{Hom}(M \oplus N, P) \cong \mathrm{Hom}(M,P) \oplus \mathrm{Hom}(N,P),$$

$$\mathrm{Hom}(M, P \oplus Q) \cong \mathrm{Hom}(M,P) \oplus \mathrm{Hom}(M,Q).$$

3. Let R be a commutative ring and let I, J be ideals in R such that $I + J = R$. Prove that $\mathrm{Hom}(R/I, R/J) = 0$.
   Conclude that the $\mathbb{R}[X]$-modules $\mathbb{R}[X]/(X-a)$ and $R[X]/(X-b)$ are not isomorphic if $a \neq b$.
4. Prove the "5-lemma": this is the following assertion. Suppose R is a ring, and given is a commutative diagram of R-modules

$$\begin{array}{ccccccccc} A & \xrightarrow{f} & B & \xrightarrow{g} & C & \xrightarrow{h} & D & \xrightarrow{j} & E \\ \downarrow \ell & & \downarrow m & & \downarrow n & & \downarrow p & & \downarrow q \\ A' & \xrightarrow{r} & B' & \xrightarrow{s} & C' & \xrightarrow{t} & D' & \xrightarrow{u} & E' \end{array}$$

   in which the two rows are exact. Show that if ℓ is surjective and q is injective and both m and p are isomorphisms, then n is an isomorphism.
5. Given a ring R and R-modules M_1, M_2, N, show that

$$(M_1 \oplus M_2) \otimes_R N \cong (M_1 \otimes_R N) \oplus (M_2 \otimes_R N).$$

6. Prove that the tensor product is right exact. This means the following. Let R be a ring and let T be an R-module. Suppose

$$M \to N \to P \to 0$$

   is an exact sequence of R-modules. Show that this results in a sequence

$$T \otimes_R M \to T \otimes_R N \to T \otimes_R P \to 0$$

   which is also exact.
7. Given the linear maps $\mathbb{R}^2 \to \mathbb{R}^2$ with matrices $\left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 5 & 6 \\ 7 & 8 \end{smallmatrix}\right)$, respectively (with respect to the standard basis of $\mathbb{R}^2$).
   Determine the Kronecker product of these linear maps (see Remark VII.3.12).
8. Let R be a commutative ring and let $I \subset R$ be an ideal.
   Prove: R/I is a projective R-module if and only if
   an ideal $J \subset R$ exists such that the canonical map $R \to R/I \times R/J$ given by $r \mapsto (r \bmod I, r \bmod J)$ defines an isomorphism of rings.
   (Hint: if g splits the canonical map $R \to R/I$, then consider $g(R/I) \subset R$.)
9. Show that the ideal $I = (X,Y) \subset \mathbb{R}[X,Y]$ is *not* projective as an R-module. (Hint: if ϕ splits the surjection

$$R \oplus R \longrightarrow I \qquad \text{given by} \quad (f,g) \mapsto fX + gY,$$

   then consider $\phi(XY) \in R \oplus R$.)
10. Suppose M is a projective module over the commutative ring R.
    Prove that $\mathrm{Hom}(M,R)$ is a projective R-module as well. (This R-module is usually called the *dual* of the R-module M).