Algebra II (Math 201)

Robert Boltje
Department of Mathematics
University of California
Santa Cruz, CA 95064
Email: boltje@math.ucsc.edu

Winter 1999

I. Modules over Principal Ideal Domains

1 Modules, Module Homomorphisms, and Basic Concepts

1.1 Remark Let M be an Abelian group (written additively).

(a) The set $\mathrm{End}(M) := \mathrm{Hom}(M, M)$ of group homomorphisms is a ring under

$$(f+g)(m) := f(m) + g(m)\,, \quad (f \cdot g)(m) := f(g(m))\,,$$

for $f, g \in \mathrm{End}(M)$ and $m \in M$. Indeed, it is very easy to verify that $f+g, f \cdot g \in \mathrm{End}(M)$ and that the ring axioms are satisfied. The 0-element and 1-element are the constant function $m \mapsto 0$ and the identity map $m \mapsto m$, respectively. $\mathrm{End}(M)$ is called the *endomorphism ring* of M. Any subring of $\mathrm{End}(M)$ is called a *ring of endomorphisms* of M.

(b) If M is generated by elements m_i, $i \in I$, and if $f, g \in \mathrm{End}(M)$ with $f(m_i) = g(m_i)$ for all $i \in I$, then $f = g$.

(c) Recall that $f(a \cdot m) = a \cdot f(m)$ for all $a \in \mathbb{Z}$, $m \in M$, and $f \in \mathrm{End}(M)$.

1.2 Example Let $M = \mathbb{Z}$. Since $\mathbb{Z}$ is generated by 1, the map

$$\alpha\colon \mathrm{End}(\mathbb{Z}) \to \mathbb{Z}\,, \quad f \mapsto f(1)\,,$$

is injective. It is also surjective, since for given $a \in \mathbb{Z}$ there is the endomorphism $\mathbb{Z} \to \mathbb{Z}$, $z \mapsto nz$. Moreover α is a ring homomorphism: For $f, g \in \mathrm{End}(\mathbb{Z})$ we have

$$\alpha(f+g) = (f+g)(1) = f(1) + g(1) = \alpha(f) + \alpha(g)$$

and

$$\begin{aligned}\alpha(f \cdot g) = (f \circ g)(1) = f(g(1)) = f(g(1) \cdot 1) = \\ g(1) \cdot f(1) = f(1) \cdot g(1) = \alpha(f) \cdot \alpha(g)\,,\end{aligned}$$

and finally $\alpha(\mathrm{id}_{\mathbb{Z}}) = \mathrm{id}_{\mathbb{Z}}(1) = 1$. There fore α is a ring isomorphism.

1.3 Remark Each ring R is isomorphic to a ring of endomorphisms on R itself. For $a \in R$ let a_l denote the left multiplication by a, so

$$a_l\colon R \to R\colon r \mapsto ar\,.$$

Then a_l is a group homomorphism, since

$$a_l(r+s) = a(r+s) = ar + as = a_l(r) + a_l(s)\,.$$

Therefore, the map

$$\alpha\colon R \to \mathrm{End}(R)\,, \quad a \mapsto a_l\,,$$

is well-defined. We will show that α is an injective ring homomorphism so that R is isomorphic to the ring $\alpha(R)$ of endomorphisms of R. Let $a, b \in R$. If $a_l = b_l$, then $a = a_l(1) = b_l(1) = b$. Moreover,

$$\begin{gathered}(\alpha(a+b))(r) = (a+b)_l(r) = (a+b)r = ar + br = \\ a_l(r) + b_l(r) = (a_l + b_l)(r) = (\alpha(a) + \alpha(b))(r)\end{gathered}$$

which implies that α is additive, and

$$(\alpha(ab))(r) = (ab)_l(r) = (ab)r = a(br) = a_l(b_l(r)) = (\alpha(a) \circ \alpha(b))(r)$$

which implies that α is multiplicative. Finally, $\alpha(1) = 1_l = \mathrm{id}_R$.

1.4 Remark Let M be an Abelian group. Then the evaluation map

$$\varphi\colon \mathrm{End}(M) \times M \to M\,, \quad (f, m) \mapsto f(m) =: f \cdot m\,,$$

satisfies

$$\begin{array}{cl} f \cdot (m+n) = f \cdot m + f \cdot n & (f \text{ is a homomorphism}) \\ (f+g) \cdot m = f \cdot m + g \cdot m & (\text{definition of } f+g) \\ (f \cdot g)(m) = f \cdot (g \cdot m)) & (\text{definition of } f \cdot g) \\ 1 \cdot m = m & (1 = \mathrm{id}_M) \end{array}$$

for all $f, g \in \mathrm{End}(M)$ and $m, n \in M$.

More generally, if R is any ring and $\rho\colon R \to \mathrm{End}(M)$ is a ring homomorphism, then the map

$$\varphi\colon R \times M\,, \quad (r, m) = (\rho(r))(m) =: r \cdot m\,,$$

satisfies

$$\begin{array}{rcl} \text{(i)} & r \cdot (m+n) = r \cdot m + r \cdot n & (\rho \text{ is a homomorphism}) \\ \text{(ii)} & (r+s) \cdot m = r \cdot m + s \cdot m & (\rho \text{ is additive}) \\ \text{(iii)} & (r \cdot s) \cdot m = r \cdot (s \cdot m) & (\rho \text{ is multiplicative}) \\ \text{(iv)} & 1_R \cdot m = m & (\rho(1_R) = \mathrm{id}_M) \end{array}$$

1.5 Definition Let R be a ring. A *left R-module* M is an Abelian group M together with a map

$$\varphi\colon R \times M \to M\,, \quad (r, m) \mapsto \varphi(r, m) =: r \cdot m\,,$$

satisfying the equations (i)–(iv) in Remark 1.4. Similarly, we define a *right* R*-module* as an Abelian group M together with a map

$$\psi\colon M \times R \to M\,, \quad (m,r) \mapsto \psi(m,r) =: m \cdot r\,,$$

satisfying

$$\begin{array}{ll} \text{(i')} & (m+n)\cdot r = m\cdot r + n\cdot r \\ \text{(ii')} & m\cdot(r+s) = m\cdot r + m\cdot s \\ \text{(iii')} & m\cdot(r\cdot s) = (m\cdot r)\cdot s \\ \text{(iv')} & m\cdot 1 = m \end{array}$$

for all $m \in M$ and $r, s \in R$. The map φ (resp. ψ) is called a *left* (resp. *right*) R*-module structure on* M. We usually write just $r\cdot m$ or rm (resp. $m\cdot r$ or mr) instead of $\varphi(r,m)$ (resp. $\psi(m,r)$).

1.6 Remark Let M be an Abelian group and let R be a ring.

(a) By Remark 1.4, each ring homomorphism $\rho\colon R \to \mathrm{End}(M)$ gives rise to a left R-module structure $\varphi\colon R\times M \to M$ on M by setting $\varphi(r,m) := (\rho(r))(m)$. Conversely, if $\varphi\colon R\times M \to M$ is a left R-module structure on M we can associate a ring homomorphism $\rho\colon R \to \mathrm{End}(M)$ by setting $(\rho(r))(m) := \varphi(r,m)$, i.e., $\rho(r)$ is just multiplication by r. These two constructions are inverse bijections between the set of left R-module structures $\varphi\colon R\times M \to M$ and the set of ring homomorphisms $\rho\colon R \to \mathrm{End}(M)$.

(b) If $\varphi\colon R\times M \to M$ is a left R-module structure on M, then

$$\varphi^\circ\colon M \times R^\circ \to M\,, \quad (m,r) \mapsto \varphi(r,m)\,,$$

is a right R°-module structure on M. (Recall that R° is the ring with the same elements and addition as in R, but with reversed multiplication: $r\cdot s$ (in R) $=$ $s\cdot r$ (in R°).) If R is commutative, we do not need to distinguish between left and right R-module structures.

(c) Similarly as in (a), the right R-module structures $\psi\colon M\times R \to M$ correspond bijectively to the ring antihomomorphisms $\sigma\colon R \to \mathrm{End}(M)$. Recall that an antihomomorphism switches the two factors of a product.

(d) If M is a left R-module, then $0\cdot m = 0 = r\cdot 0$ for all $m \in M$ and all $r \in R$, since $0\cdot m = (0+0)\cdot m = 0\cdot m + 0\cdot m$ and $r\cdot 0 = r\cdot(0+0) = r\cdot 0 + r\cdot 0$. Moreover, it is easy to show that $(-r)\cdot m = -(r\cdot m) = r\cdot(-m)$ for all $r\in R$ and $m \in M$.

(e) If R is a field, then an R-module is the same as an R-vector space.

(f) When we say 'R-module' in the sequel, we will always mean 'left R-module'.

1.7 Examples (a) Let R be a ring. Then the ring multiplication $R\times R \to R$, $(r,s)\mapsto rs$, defines a left and also a right R-module structure on R. We denote these two modules by ${}_RR$ and R_R. More generally, if $S \subseteq R$ is a subring (this always means with the same 1-element), then R is a left and right S-module just by the multiplication defined inside R.

(b) Each Abelian group is a left module over its endomorphism ring.

(c) Each Abelian group is automatically a $\mathbb{Z}$-module in a unique way. Therefore, $\mathbb{Z}$-modules and Abelian groups are the same mathematical notions.

(d) Let F be a field and let V be an F-vector space. For each F-linear map $f\colon V \to V$ and each polynomial

$$p(\lambda) = a_n\lambda^n + \cdots + a_1\lambda + a_0 \in F[\lambda]$$

we set

$$p(f) := a_n f^n + \cdots + a_1 f + a_0 \cdot \mathrm{id}_V \,.$$

Obviously, $p(f)$ is again an F-linear map from V to V and we obtain for fixed f a ring homomorphism

$$F[\lambda] \to \mathrm{End}_F(V) \subseteq \mathrm{End}(V)\,, \quad p \mapsto p(f)\,.$$

Therefore, V becomes an $F[\lambda]$-module (whose structure depends on f) by $p \cdot v = (p(f))(v)$, for $p \in F[\lambda]$ and $v \in V$.

1.8 Definition Let M and N be R-modules. A map $f\colon M \to N$ is called an *R-module homomorphism* (or R-linear) if it is a group homomorphism and respects the R-module structure, i.e., $f(rm) = rf(m)$ for all $r \in R$ and $m \in M$. We denote the set of R-module homomorphisms $f\colon M \to N$ by $\mathrm{Hom}_R(M,N)$. Note that $\mathrm{Hom}_R(M,N)$ is a subgroup of $\mathrm{Hom}(M,N)$ and that the composition of module homomorphisms is again a module homomorphism. Similarly, we define homomorphisms for right R-modules. As for groups we define the notions of *monomorphisms, epimorphisms, isomorphisms, endomorphisms*, and *automorphisms* of modules in the obvious way. We call R-modules M and N *isomorphic* (and write $M \cong N$), if there exists an isomorphism between M and N. Note that if $f\colon M \to N$ is a bijective R-linear map, then also f^{-1} is R-linear. Note also that if $R = \mathbb{Z}$, then an R-linear map is just a group homomorphism.

1.9 Definition Let R be a ring and let M be an R-module. An *R-submodule* of M is a subgroup U of M such that $ru \in U$ for all $r \in R$ and $u \in U$.

1.10 Examples Let R be a ring and let M be an R-module.

(a) The subgroups $\{0\}$ and M of M are always R-submodules. If these are the only submodules and if $M \neq \{0\}$, then we call M an *irreducible* (or *simple*) R-module.

(b) Let N be an R-module and let $f \in \mathrm{Hom}_R(M,N)$. Then the *image* $\mathrm{im}(f) = f(M)$ of f is an R-submodule of N. More generally, the image $f(U)$ of any R-submodule U of M is again an R-submodule of N. Similarly, the *kernel* $\ker(f)$ is an R-submodule of M. More generally, the preimage $f^{-1}(V)$ of any R-submodule V of N is again an R-submodule of M.

(c) If U is an R-submodule of M, then the factor group M/U carries again a natural R-module structure which is defined by

$$r \cdot (m+U) := r \cdot m + U$$

for $r \in R$, $m \in M$. This is well-defined, since for $m' \in M$ with $m - m' \in U$ we have $r \cdot m - r \cdot m' = r \cdot (m - m') \in U$ and therefore $r \cdot m + U = r \cdot m' + U$. The map

$$\nu \colon M \to M/U\,, \quad m \mapsto m + U\,,$$

is R-linear and is called the *natural epimorphism*.

(d) Let $X \subseteq M$ be a subset of M. Then, the set

$$\langle X \rangle_R := \{r_1 x_1 + \cdots + r_n x_n \mid n \in \mathbb{N}_0,\ r_1, \ldots, r_n \in R,\ x_1, \ldots, x_n \in X\}$$

is an R-submodule of M and is called the R-submodule of M *generated by* X. It is easy to see that this is the smallest R-submodule of M containing X, i.e.,

$$\langle X \rangle_R = \bigcap_{\substack{X \subseteq U \subseteq M \\ U \text{ is } R\text{-submodule of } M}} U\,.$$

We call M a *finitely generated* R-module, if M is generated by a finite set. We call M *cyclic*, if $M = \langle m \rangle_R = Rm$ for some $m \in M$.

(e) If U_i, $i \in I$, are R-submodules of M, then also their intersection $\bigcap_{i \in I} U_i$ is an R-submodule of M. Moreover, we define the *sum* $\sum_{i \in I} U_i$ of U_i, $i \in I$, as the R-submodule generated by $\bigcup_{i \in I} U_i$. This is the smallest R-submodule of M containing every U_i, $i \in I$.

(f) If M_i, $i \in M$, are R-modules, then their *direct product* $\prod_{i \in I} M_i$ is again an R-module with the componentwise group structure and the R-module structure given by $r \cdot (m_i)_{i \in I} := (r \cdot m_i)_{i \in I}$. Moreover, the *external direct sum* $\bigoplus_{i \in I} M_i$ is defined as the R-submodule of $\prod_{i \in I} M_i$ consisting of those families $(m_i)_{i \in I}$ which have only finitely many non-zero entries. If I is finite, then $\prod_{i \in I} M_i = \bigoplus_{i \in I} M_i$. We abbreviate the direct sum of n copies of ${}_R R$ by $R^{(n)}$.

(g) The R-submodules of ${}_R R$ are just the *left ideals* and those of R_R are the *right ideals* of R.

(h) For any subset X of M, the *annihilator*

$$\mathrm{ann}_R(X) := \{r \in R \mid rx = 0 \text{ for all } x \in X\}$$

of X is a left ideal of R. If U is an R-submodule of M, then $\mathrm{ann}_R(U)$ is a (two-sided) ideal of R.

1.11 Theorem (Fundamental theorem of homomorphisms for modules) *Let M and N be R-modules and let $f \in \mathrm{Hom}_R(M, N)$. For each R-submodule U of M with $U \subseteq \ker(f)$, there exists a unique $\bar{f} \in \mathrm{Hom}_R(M/U, N)$ such that the diagram*

$$\begin{tikzcd}
M \arrow[r, "f"] \arrow[d, "\nu"'] & N \\
M/U \arrow[ur, "\bar{f}"'] &
\end{tikzcd}$$

commutes, where $\nu\colon M \to M/U$ *is the natural epimorphism. Moreover,* $\bar{f}$ *has the same image as* f *and* $\ker(\bar{f}) = \ker(f)/U$*. The map*

$$\mathrm{Hom}_R(M/U, N) \to \{f \in \mathrm{Hom}_R(M,N) \mid f(U) = 0\}\,, \quad g \mapsto g \circ \nu\,,$$

is a group isomorphism.

Proof Everything follows from the fundamental theorem of group homomorphisms; except that one has to verify that the group homomorphism $\bar{f}$ is R-linear. But, for $r \in R$ and $m \in M$, we have

$$\begin{aligned}
\bar{f}(r(m+U)) = \bar{f}(rm+U) = (\bar{f}\circ\nu)(rm) = \\
f(rm) = rf(m) = r\bar{f}(\nu(m)) = r\bar{f}(m+U)\,,
\end{aligned}$$

and the proof is complete. □

1.12 Remark If R is a ring, M is an R-module, and U is an R-submodule, then the natural epimorphism $\nu\colon M \to M/U$ induces a bijection $V \mapsto V/U$ between the set of R-submodules of M containing U and the set of all R-submodules of M/U. Moreover, the isomorphism theorems hold similar as for groups, i.e., there are natural R-linear isomorphisms

$$(M/U)/(V/U) \cong M/V \quad \text{and} \quad (W+U)/U \cong W/(W \cap U)$$

for all R-submodules $U \subseteq V \subseteq M$ and $W \subseteq M$.

1.13 Definition Let M be an R-module. A subset X of M is called an R-*basis* of M, if $\langle X \rangle_R = M$ and if X is an R-*linearly independent* set, i.e., any R-linear combination of finitely many pairwise distinct elements in X which equals 0 has 0-coefficients. We call M a *free* R-module, if M has a basis. We call M a *finitely generated free* R-module, if M has an R-basis of finite length. Note that this is the same as saying that M is free and finitely generated. Free $\mathbb{Z}$-modules are also called *free Abelian groups.*

1.14 Remark Let R be a ring and let M be an R-module.

(a) A subset $X \subseteq M$ is an R-basis of M if and only if each element of M can be written in a unique way as an R-linear combination of pairwise distinct elements of X.

(b) For each $n \in \mathbb{N}$, the R-module $R^{(n)}$ is finitely generated free with R-basis $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, the 1 being at the i-th position, $i = 1, \ldots, n$.

1.15 Theorem (Universal property of an R-basis) *Let M be a free R-module with R-basis X. Let N be another R-module and let $g\colon X \to N$ be a function. Then there exists a unique homomorphism $f \in \mathrm{Hom}_R(M,N)$ such that $f|_X = g$.*

Proof Existence: We write each $m \in M$ as R-linear combination $m = \sum_{x\in X} r_x x$ with unique coefficients $r_x \in R$, $x \in X$, all but finitely many being 0. For m as above we set $f(m) := \sum_{x\in X} r_x g(x)$. Then, for another element $n = \sum_{x\in X} s_x x$ of M and for $r \in R$, we have $m + n = \sum_{x\in X}(r_x + s_x)x$ and $rm = \sum_{x\in X}(rr_x)x$. Hence,

$$f(m+n) = \sum_{x\in X}(r_x + s_x)g(x) = \sum_{x\in X} r_x g(x) + \sum_{x\in X} s_x g(x) = f(m) + f(n)$$

and

$$f(rm) = \sum_{x\in X}(rr_x)g(x) = r\sum_{x\in X} r_x g(x) = rf(m)\,.$$

Therefore, f is R-linear, and obviously $f|_X = g$.

The uniqueness of f is clear, since X generates M as R-module. □

1.16 Corollary *Any finitely generated free R-module is isomorphic to $R^{(n)}$ for some $n \in \mathbb{N}_0$.*

Proof Let $X = \{x_1, \ldots, x_n\}$ be an R-basis of M. Let $e_1, \ldots, e_n$ be the canonical R-basis of $R^{(n)}$. Then by the universal property of R-bases there exist $f_1 \in \mathrm{Hom}_R(M, R^{(n)})$ and $f_2 \in \mathrm{Hom}_R(R^{(n)}, M)$ such that $f_1(x_i) = e_i$ and $f_2(e_i) = x_i$ for all $i = 1, \ldots, n$. This implies $f_2 \circ f_1 = \mathrm{id}_M$ and $f_1 \circ f_2 = \mathrm{id}_{R^{(n)}}$, so that f_1 and f_2 are inverse isomorphisms. □

1.17 Example Is $x_1 = (1,2)$, $x_2 = (2,1)$ a $\mathbb{Z}$-basis of $\mathbb{Z}^{(2)}$? We check linear independence: If $a(1,2)+b(2,1) = 0$ for $a, b \in \mathbb{Z}$, then $a+2b = 0$ and $2a+b = 0$. This implies $a = -2b$ and substitution in the second equation gives $b = 0$ and then we also have $a = 0$. Thus x_1, x_2 are $\mathbb{Z}$-linearly independent. In order to show that $\langle x_1, x_2\rangle_R = \mathbb{Z}^{(2)}$ we have to verify that $(1,0), (0,1) \in \langle x_1, x_2\rangle_R$. But this turns out to be impossible: Suppose $(1,0) = a(1,2) + b(2,1)$ with $a, b \in \mathbb{Z}$. Then $1 = a + 2b$ and $0 = 2a + b$. So $b = -2a$ and substituting this in the first equation we obtain $1 = -3a$ which is not possible for $a \in \mathbb{Z}$. Thus, x_1, x_2 is not a $\mathbb{Z}$-basis of $\mathbb{Z}^{(2)}$.

1.18 Theorem *Let R be a commutative ring and let M be a finitely generated free R-module. Then any two R-bases of M have the same length. This length is called the* R-rank *of M and is denoted by $\mathrm{rk}_R(M)$.*

Proof Let $x_1, \dots, x_m$ and $y_1, \dots, y_n$ be R-bases of M and assume that $m < n$. We can write

$$y_i = \sum_{j=1}^{m} a_{ji} x_j \quad \text{and} \quad x_k = \sum_{l=1}^{n} b_{lk} y_l$$

for all $i = 1, \dots, n$ and all $k = 1, \dots, m$ with elements $a_{ji}, b_{lk} \in R$ for $1 \leqslant j, k \leqslant m$ and $1 \leqslant i, l \leqslant n$. Substitution yields

$$y_i = \sum_{j=1}^{m} a_{ji} \sum_{l=1}^{n} b_{lj} y_l = \sum_{j=1}^{m} \sum_{l=1}^{n} a_{ji} b_{lj} y_l = \sum_{l=1}^{n} (\sum_{j=1}^{m} b_{lj} a_{ji}) y_l$$

for all $i = 1, \dots, n$. Since $y_1, \dots, y_n$ are R-linearly independent, we obtain

$$\sum_{j=1}^{m} b_{lj} a_{ji} = \delta_{li}$$

for all $1 \leqslant i, l \leqslant n$. The $n \times n$-matrices

$$A := \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \quad \text{and } B := \begin{pmatrix} b_{11} & \cdots & b_{1m} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nm} & 0 & \cdots & 0 \end{pmatrix}$$

then satisfy the relation $BA = 1_n$. Therefore, $\det(B)\det(A) = 1$. On the other hand, obviously $\det(A) = 0$. This is a contradiction. □

1.19 Theorem (Universal property of direct sums) *Let M_i, $i \in I$, be R-modules and let $M = \bigoplus_{i \in I} M_i$. We denote by $\iota_i \colon M_i \to M$ the canonical embedding which maps $m_i \in M_i$ to the family which has the entry m_i at i and entries 0 everywhere else. Let also N be an R-module. Then, for each family of homomorphisms $f_i \in \operatorname{Hom}_R(M_i, N)$, $i \in I$, there exists a unique homomorphism $f \in \operatorname{Hom}_R(M, N)$ such that $f_i = f \circ \iota_i$ for each $i \in I$.*

Proof Let $m := (m_i)_{i \in I} \in \bigoplus_{i \in I} M_i$. We set $f(m) := \sum_{i \in I} f_i(m_i)$. Then it is easy to verify that $f \in \operatorname{Hom}_R(M, N)$ and that $f_i = f \circ \iota_i$ for each $i \in I$. Moreover, f is unique with this property, since M is the sum of the submodules $\iota_i(M_i)$. □

1.20 Definition Let M be an R-module and let $M_i, i \in I$, be submodules of M. M is called the *internal direct sum*, if the unique homomorphism $f \colon \bigoplus_{i \in I} M_i \to M$, such that $f \circ \iota_i$ is the embedding $M_i \subseteq M$, is an isomorphism. In this case we also write $M = \bigoplus_{i \in I} M_i$ and identify M with the external direct sum via f. Note that, in this case, if $g_i \colon M_i \to N$ are R-linear maps to some R-module N, then there exists a unique R-linear map $g \colon M \to N$ such that $g|_{M_i} = g_i$.

1.21 Theorem *Let R be a ring, let M be an R-module, and let U_i, $i \in I$, be R- submodules of M. Then the following are equivalent:*

(i) *M is the internal direct sum of the U_i, $i \in I$.*

(ii) (α) *$M = \sum_{i \in I} U_i$, and*

(β) *$U_{i_0} \cap \sum_{i \in I \smallsetminus \{i_0\}} U_i = 0$ for all $i_0 \in I$.*

(iii) *Each element $m \in M$ can be written as $m = \sum_{i \in I} u_i$ with unique elements $u_i \in U_i$, $i \in I$ (only finitely many of which are non-zero).*

Proof (i) $\Rightarrow$ (ii): If $f \colon \bigoplus_{i \in I} U_i \to M$ as in Definition 1.20 is an isomorphism, then $\iota_i(U_i) \subseteq \bigoplus_{i \in I} U_i$ and $U_i \subseteq M$ correspond under f. Since $\sum_{i \in I} \iota_i(U_i) = \bigoplus_{i \in I} U_i$, ($\alpha$) holds, and since $\iota_{i_0}(U_{i_0}) \cap \sum_{i \in I \smallsetminus \{i_0\}} \iota_i(U_i) = 0$ for all $i_0 \in I$, (β) holds.

(ii) $\Rightarrow$ (iii): By (α), each $m \in M$ can be written as $m = \sum_{i \in I} u_i$ with $u_i \in U_i$ for $i \in I$. By (β), the elements $u_i \in U_i$ are uniquely determined by m. In fact, if also $v_i \in U_i$, $i \in I$, satisfy $m = \sum_{i \in I} v_i$, and if $u_{i_0} \neq v_{i_0}$ for some $i_0 \in I$, then subtraction yields $0 \neq v_{i_0} - u_{i_0} = \sum_{i \in I \smallsetminus \{i_0\}} u_i - v_i \in U_{i_0} \cap \sum_{i \in I \smallsetminus \{i_0\}} U_i$.

(iii) $\Rightarrow$ (i): The map $f \colon \bigoplus_{i \in I} U_i \to M$ defined on the external direct sum $\bigoplus_{i \in I} U_i$ as in Definition 1.20 is surjective, since each $m \in M$ can be expressed as sum of elements in U_i, $i \in I$. Moreover, the uniqueness of such an expression implies the injectivity of f. $\square$

1.22 Examples (a) Consider the $\mathbb{Z}$-submodules $U = \{(a, b, 0) \mid a, b \in \mathbb{Z}\}$ and $V = \{(c, c, c) \mid c \in \mathbb{Z}\}$ of $M = \mathbb{Z}^{(3)}$. Then $M = U \oplus V$. In fact, each element $(a, b, c) \in M$ can be written as $(a, b, c) = (a - c, b - c, 0) + (c, c, c) \in U + V$, and if $(a, b, 0) = (c, c, c)$ for integers a, b, c, then $a = c$, $b = c$, and $0 = c$ imply $a = b = c = 0$. But note that M is not the internal direct sum of U and $2V = \{(2c, 2c, 2c) \mid c \in \mathbb{Z}\}$, since $(0, 0, 1) \notin U + 2V$.

(b) The $\mathbb{Z}$-module $M = \mathbb{Z}/6\mathbb{Z}$ is the internal direct sum of the submodules $U = 3\mathbb{Z}/6\mathbb{Z} = \{0 + 6\mathbb{Z}, 3 + 6\mathbb{Z}\}$ and $V = 2\mathbb{Z}/6\mathbb{Z} = \{0 + 6\mathbb{Z}, 2 + 6\mathbb{Z}, 4 + 6\mathbb{Z}\}$ by Theorem 1.21 (iii). (This can also be seen by applying the Chinese Remainder Theorem.)

(c) Let V be a vector space over a field F and let $f \in \mathrm{End}_F(V)$. Consider the $F[\lambda]$-module structure of V given by extending $\lambda \cdot v := f(v)$ to $F[\lambda]$, cf. Example 1.7 (d). The $F[\lambda]$-submodules of V are precisely those F-subspaces U of V with $f(U) \subseteq U$. We call such a subspace *f-stable*.

(d) As a more concrete example, let V be a 3-dimensional vector space over $\mathbb{R}$ with $\mathbb{R}$-basis x, y, z, and let $f \in \mathrm{End}_{\mathbb{R}}(V)$ be given by $f(x) = 3x$, $f(y) = y + z$, and $f(z) = -y + 2z$. Then the subspaces $U = \mathbb{R}x$ and $V = \mathbb{R}y + \mathbb{R}z$ are f-stable. Therefore, with the $F[\lambda]$-module structure of V defined as in (c), we have an internal direct sum decomposition $M = U \oplus V$ of M into $F[\lambda]$-submodules.

1.23 Remark (Homomorphisms between free modules and matrices) Throughout this remark let R be a commutative ring.

(a) Let M be an R-module. Since R is commutative we may consider M as left and as right R-module. Using these module structures we can define the matrix product

$$A \cdot Q \quad \text{and} \quad Q \cdot B$$

for any matrices A and B with entries in R and any matrix Q with entries in M, as long as their sizes are compatible for multiplication. The usual rules for matrix addition and multiplication still hold in this situation.

(b) Let M and N be free R-modules with ordered R-bases $X = (x_1, \dots, x_m)$ and $Y = (y_1, \dots, y_n)$, respectively. Let $f \in \mathrm{Hom}_R(M, N)$ and, for $i = 1, \dots, m$, write

$$f(x_i) = \sum_{j=1}^{n} a_{ji} y_j \tag{1.23.a}$$

with $a_{ji} \in R$ for $1 \leqslant j \leqslant n$, $1 \leqslant i \leqslant m$. Then, f is uniquely determined by the elements $a_{ji} \in R$, and conversely, for any given set of elements $a_{ji} \in R$, $1 \leqslant j \leqslant n$, $1 \leqslant i \leqslant m$, we may define an R-linear map $f \colon M \to N$ by the Equations (1.23.a) using the universal property of a basis. If we arrange the elements a_{ji} in the form of an $n \times m$-matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix},$$

then the above arguments show that we obtain a bijection between $\mathrm{Hom}_R(M, N)$ and $M_{n\times m}(R)$. This is a group isomorphisms, since if the matrix B is attached to a further map $g \in \mathrm{Hom}_R(M, N)$ then the matrix $A + B$ is attached to $f + g$. We say that *A represents f with respect to the bases X and Y*, and we denote A by ${}_Y M_X(f)$. The i-th column of ${}_Y M_X(f)$ gives the coefficients of $f(x_i)$ with respect to the basis Y. If we write the elements $R^{(m)}$ and $R^{(n)}$ as columns and identify them with M and N, respectively, by

$$\begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} \mapsto \sum_{i=1}^{m} r_i x_i \quad \text{and} \quad \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \mapsto \sum_{j=1}^{n} r_j y_j\,,$$

then f translates to an R-linear map $R^{(m)} \to R^{(n)}$ which is given as the multiplication by A:

$$\begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} \mapsto A \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}.$$

Note that with the notation in (a) we can also write the Equations (1.23.a) in the form

$$(f(x_1), \dots, f(x_m)) = (y_1, \dots, y_n)A\,.$$

(c) If P is a further free R-module with finite R-basis $Z = (z_1, \dots, z_p)$ and if $g \in \mathrm{Hom}_R(N, P)$, then we have

$$_Z M_X(g \circ f) = {}_Z M_Y(g) \cdot {}_Y M_X(f) .$$

In fact, if $_Z M_Y(g) = (g_{kj})$, then the i-th column of $_Z M_X(g \circ f)$ is determined by

$$g(f(x_i)) = \sum_{j=1}^{n} a_{ji} g(y_j) = \sum_{j=1}^{n} a_{ji} \sum_{k=1}^{p} b_{kj} z_k = \sum_{k=1}^{p} \Big(\sum_{j=1}^{n} b_{kj} a_{ji} \Big) z_k ,$$

and the coefficient of z_k is precisely the (k, i)-entry of $B \cdot A$.

(d) If we use other R-bases $X' = (x'_1, \dots, x'_m)$ and $Y' = (y'_1, \dots, y'_n)$ for M and N, respectively, then by (c) we have

$$_{Y'} M_{X'}(f) = {}_{Y'} M_Y(\mathrm{id}_N) \cdot {}_Y M_X(f) \cdot {}_X M_{X'}(\mathrm{id}_M) .$$

More explicitly, if

$$\mathrm{id}_M(x'_i) = x'_i = \sum_{j=1}^{m} s_{ji} x_j \quad \text{and} \quad \mathrm{id}_N(y'_k) = y'_k = \sum_{l=1}^{n} t_{lk} y_l$$

and if we define the matrices $S = (s_{ji})$ and $T = (t_{lk})$, then $S = {}_X M_{X'}(\mathrm{id}_M)$ and $T = {}_Y M_{Y'}(\mathrm{id}_N)$. These matrices are invertible, since by part (c) we have

$$_X M_{X'}(\mathrm{id}_M) \cdot {}_{X'} M_X(\mathrm{id}_M) = {}_X M_X(\mathrm{id}_M) = 1_m ,$$

and similar in the reversed order and for Y and Y'. This also shows that $S^{-1} = {}_{X'} M_X(\mathrm{id}_M)$ and $T^{-1} = {}_{Y'} M_Y(\mathrm{id}_N)$ so that

$$_{Y'} M_{X'}(f) = T^{-1} A S .$$

(e) If $x'_1, \dots, x'_m$ are arbitrary elements of M and if we write

$$x'_i = \sum_{j=1}^{m} s_{ji} \cdot x_j ,$$

for each $i = 1, \dots, m$, then the matrix $S = (s_{ji})$ is invertible in $M_m(R)$ if and only if $x'_1, \dots, x'_m$ is again an R-basis of M. In fact, if $x'_1, \dots, x'_m$ is an R-basis, then S is invertible by part (d). Conversely, assume that S is invertible. The above equations can be written as

$$(x'_1, \dots, x'_m) = (x_1, \dots, x_m) S ,$$

so that we have

$$(x_1, \dots, x_m) = (x'_1, \dots, x'_m) S^{-1} .$$

This shows that $x_1, \ldots, x_m \in \langle x'_1, \ldots x'_m \rangle_R$ and therefore, $M = \langle x'_1, \ldots, x'_m \rangle_R$. Also, if $r_1 x'_1 + \cdots + r_m x'_m = 0$ for elements $a_1, \ldots, a_m \in R$, then

$$(r_1, \ldots, r_m) \begin{pmatrix} x'_1 \\ \vdots \\ x'_m \end{pmatrix} = 0$$

and so

$$(r_1, \ldots, r_m) S^t \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = 0 \,.$$

Since $x_1, \ldots, x_m$ are linearly independent, we obtain $(r_1, \ldots, r_m)S^t = 0$, and, since S^t is invertible, also $(r_1, \ldots, r_m) = 0$.

Recall that, for commutative R, a square matrix S with coefficients in R is invertible if and only if $\det(S)$ is invertible in R, cf. [J, Theorem 2.1, p. 96].

1.24 Examples (a) Let V be an $\mathbb{R}$-vector space with basis x, y, z and let $f \in \mathrm{End}_{\mathbb{R}}(V)$ be given by $f(x) = 3x$, $f(y) = y + z$, $f(z) = -y + 2z$ as in Example 1.22 (c). Then the matrix representing f with respect to the basis (x, y, z) is

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 2 \end{pmatrix} \,.$$

If we want to represent f with respect to the basis $(x, y, y + z)$, we have to find $S \in \mathrm{GL}_3(\mathbb{R})$ such that

$$(x, y, y + z) = (x, y, z)S \,.$$

This implies

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

so that f is represented with respect to the basis $(x, y, y + z)$ by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} =$$
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & -3 \\ 0 & 1 & 3 \end{pmatrix} \,.$$

(b) Consider the elements $(0,1,2), (2,3,1), (-1,-1,0) \in \mathbb{Z}^{(3)}$. Do they form a $\mathbb{Z}$-basis of $\mathbb{Z}^{(3)}$? We have to determine, if $\det(S)$ is invertible in $\mathbb{Z}$ for

$$S = \begin{pmatrix} 0 & 2 & -1 \\ 1 & 3 & -1 \\ 2 & 1 & 0 \end{pmatrix} .$$

Now, $\det(S) = -4 - 1 + 6 = 1$. Therefore, the above elements form a $\mathbb{Z}$-basis of $\mathbb{Z}^{(3)}$.

2 Modules over Principal Ideal Domains

2.1 Definition Let R be a domain (i.e., a commutative ring without zero divisors except 0 and with $1 \neq 0$) and let M be an R-module. An element $m \in M$ is called an *R-torsion element*, if there exists $r \in R \smallsetminus \{0\}$ such that $rm = 0$. The set $\operatorname{tor}_R(M)$ of torsion elements is an R-submodule of M, called the *R-torsion submodule* of M. M is called *R-torsion free*, if $\operatorname{tor}_R(M) = 0$. Note that R-linear maps map R-torsion elements to R-torsion elements, and that a free R-module is R-torsion free. If $R = \mathbb{Z}$, we just write $\operatorname{tor}(M)$ and call this the *torsion subgroup* of M. Obviously, $\operatorname{tor}(M)$ consists of the elements of finite order.

2.2 Theorem *Let M be a finitely generated free R-module over a principal ideal domain R and let U be an R-submodule of M. Then also U is R-free and $\operatorname{rk}_R(U) \leqslant \operatorname{rk}_R(M)$.*

Proof By induction on $n = \operatorname{rk}_R(M)$. If $n = 0$, then $M = 0$, $U = 0$, and U is R-free with basis $\emptyset$. Let $n > 0$. Choose an R-basis $f_1, \ldots, f_n$ of M and set $M' := \langle f_2, \ldots, f_n \rangle_R \subseteq M$. Then M is R-free of rank $n-1$. If U is contained in M', then by induction U is R-free with $\operatorname{rk}_R(U) \leqslant n-1$. So we can assume that $U \not\subseteq M'$. Consider the R-linear map $\pi \colon M \to R$, $r_1 f_1 + \cdots + r_n f_n \mapsto r_1$. Since R is a principal ideal domain and $U \not\subseteq M'$, there exists $0 \neq d \in R$ with $\pi(U) = dR$. Let $u_1 \in U$ with $\pi(u_1) = d$. We claim that $U = Ru_1 \oplus (M' \cap U)$. In fact, $Ru_1 + (U \cap M') \subseteq U$ and $Ru_1 \cap U \cap M' \subseteq Ru_1 \cap M' = 0$. Moreover, let $u \in U$ be arbitrary and let $\pi(u) = rd$ with $r \in R$. Then $u = ru_1 + (u - ru_1)$ with $\pi(u - ru_1) = 0$, so $u - ru_1 \in U \cap M'$. By induction, $U \cap M'$ is R-free of rank at most $n-1$. Let $e_2, \ldots, e_k$ be an R-basis of $U \cap M'$ with $1 \leqslant k \leqslant n$. Then obviously $e_1 := u_1, e_2, \ldots, e_k$ is an R-basis of $U = Ru_1 \oplus (U \cap M')$. □

2.3 Definition Let R be a commutative ring. Tow matrices $A, B \in M_{m \times n}(R)$ are called *R-equivalent*, if there exist $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$. This defines an equivalence relation on $M_{m \times n}(R)$.

2.4 Remark Let R be a principal ideal domain.

(a) Let $a_1, \ldots, a_n$ be elements in R. Recall that an element $d \in R$ is called a greatest common divisor (or short a gcd) of $a_1, \ldots, a_n$ if d divides each a_i

$(1 \leqslant i \leqslant n)$ and if for any other element $e \in R$ which also divides each a_i $(1 \leqslant i \leqslant n)$ one has $e \mid d$. Recall that d is a gcd of $a_1, \dots, a_n$ if and only if $Ra_1 + \cdots Ra_n = dR$. Therefore, two gcds d, d' of $a_1, \dots, a_n$ are always associates in R, i.e., there exists a unit u of R such that $d' = du$.

(b) Let $A_1 \in M_{m\times n}(R)$ and $A_2 \in M_{n\times p}(R)$. Recall that, for $i \in \mathbb{N}$, an i-minor of A_1 is the determinant of an $i \times i$-submatrix of A arising from choosing any i rows and i columns of A_1. For $i \in \mathbb{N}$, let $\Delta_i(A_1)$ be a gcd of the i-minors of A_1. If i is bigger than m or n, then we set $\Delta_i(A_1) = 0$. Similarly we define $\Delta_i(A_2)$ and $\Delta_i(A_1A_2)$. Since the columns of A_1A_2 are R-linear combinations of those of A_1, also the columns of any $(i \times i)$-submatrix S of A_1A_2 are R-linear combinations of i-subcolumns of A_1 (using the same i rows). Since

$$\begin{gathered}\det(x_1, \dots, x_{j-1}, rx_j + sy_j, x_{j+1}, \dots, x_i) = \\ r\det(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_i) + s\det(x_1, \dots, x_{j_1}, y_j, x_{j+1}, \dots, x_i)\,,\end{gathered}$$

for any columns $x_1, \dots, x_n, y_j$ with i entries in R $(1 \leqslant j \leqslant i)$ and any $r, s \in R$, we can express $\det(S)$ as R-linear combination of i-minors of A_1. Therefore, $\Delta(A_1) \mid \Delta(A_1A_2)$. Similarly, using row arguments, we have $\Delta_i(A_2) \mid \Delta(A_1A_2)$. In particular, if $A, B \in M_{m\times n}(R)$ are R-equivalent, then $\Delta_i(A)$ and $\Delta_i(B)$ are associates in R. Note that the 1-minors of A are just the entries of A.

(c) Consider the $(n \times n)$-matrices

$$T_{ij}(a) = \begin{pmatrix} 1 & & & \\ & \ddots & a & \\ & & \ddots & \\ & & & 1 \end{pmatrix},$$

with $a \in R$ in the i-th row and j-th column $1 \leqslant i, j \leqslant n$, $i \neq j$,

$$D_i(u) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & u & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix},$$

with a unit u of R, and

$$P_{ij} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix},$$

with $1 \leqslant i, j \leqslant n$, $i \neq j$. The diagonal entries of P_{ij} are all equal to 1 except the two 0-entries at the i-th and j-th position. Multiplication by $T_{ij}(a)$ from the left (resp. right) to a matrix A adds the j-th row (resp. i-th column) of A multiplied by a to the i-th row (resp. j-th column) of A. Multiplication by $D_i(u)$ from the left (resp. right) multiplies the i-th row (resp. column) of A by u. Multiplication by P_{ij} from the left (resp. right) interchanges the i-th and j-th row (resp. column) of A. Obviously,

$$\begin{aligned} T_{ij}(a)T_{ij}(-a) &= 1_n \,, \\ D_i(u)D_i(u^{-1}) &= 1_n \,, \\ P_{ij}P_{ij} &= 1_n \,, \end{aligned}$$

so that all these matrices lie in $\mathrm{GL}_n(R)$.

2.5 Theorem *Let R be a principal ideal domain. Then each matrix $A = (a_{ij}) \in M_{m\times n}(R)$ is equivalent to a matrix of the form*

$$\begin{pmatrix} d_1 & & & & \\ & \ddots & & & \\ & & d_k & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}$$

with non-zero elements $d_1, \ldots, d_k \in R$ such that $d_1 \mid d_2 \mid \cdots \mid d_k$. Moreover, k is uniquely determined by A and $d_1, \ldots, d_k$ are unique up to multiplication with units of R. The elements $d_1, \ldots, d_k$ are called invariant factors *of A.*

Proof We define the *length* $l(r)$ of an element $r \in R$ as the number of factors occurring in a prime decomposition of R. If R is a unit we set $l(r) = 0$, and for $r = 0$ we set $l(0) = \infty$. We will first show that A is equivalent to a matrix of the above form. If $A = 0$ or if A is a 1×1-matrix, this is obvious. So we assume that $A \neq 0$ and that A has at least two entries.

Step 1: We show that A is similar to a matrix of the form

$$B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & C & \\ 0 & & & \end{pmatrix} \tag{2.5.a}$$

with an $(m-1)\times(n-1)$-matrix $C = (c_{ij})$ such that $b_{11} \mid c_{ij}$ for all $1 \leqslant i \leqslant m-1$, $1 \leqslant j \leqslant n-1$. We will prove this by induction on $l(A) := \min\{l(a_{ij}) \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$.

If $l(A) = 0$, then there is an element a_{ij} in A which is a unit. Using permutations of rows and columns we may assume that a_{11} is a unit. Then we can add the first row multiplied by $-a_{i1}a_{11}^{-1}$ to the i-th row $(i = 2, \ldots, m)$ and the first column multiplied by $-a_{1j}a_{11}^{-1}$ to the j-th column ($j = 2, \dots, n$) and obtain the desired matrix B.

If $l > 0$, then we can again assume that $l(a_{11})$ is minimal among the $l(a_{ij})$. If every other entry of A is divisible by a_{11}, we set $b_{11} = a_{11}$ and can add appropriate multiples of the first row and column to the other rows and columns to obtain the desired matrix B. Note that all the entries which are not in the first row or first column remain divisible by $a_{11} = b_{11}$, since a_{11} is a gcd of the 1-minors of A and so of B.

If there exists an entry a_{ij} in A which is not divisible by a_{11}, then we may assume that $a_{ij} = a_{12}$ by permutations of rows and columns if necessary (or $a_{ij} = a_{21}$ if there is only a single column; but this case is handled similarly). We set $a := a_{11}$, $b := a_{12}$, and choose a gcd d of a and b. Then $l(d) < l(a_{11})$. Moreover, there exist $x, y \in R$ with $ax + by = d$ and we can write $a = rd$ and $b = sd$ with $r, s \in R$. We set

$$Q := \begin{pmatrix} x & s & 0 & \cdots & 0 \\ y & -r & 0 & \cdots & \vdots \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in M_n(R)\,.$$

Note that Q is invertible, since

$$\begin{pmatrix} x & s \\ y & -r \end{pmatrix} \begin{pmatrix} r & s \\ y & -x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} r & s \\ y & -x \end{pmatrix} \begin{pmatrix} x & s \\ y & -r \end{pmatrix}\,.$$

Now the first row of AQ has the form $(d, 0, *, \dots, *)$, and since $l(d) < l(a_{11}) = l(A)$, we can now apply induction to this matrix.

Step 2: We next prove the assertion by induction on $\min\{m, n\}$. If $m = 1$ or $n = 1$, we are finished by Step 1. So we assume that $m, n \geqslant 2$. Then, by Step 1, A is equivalent to a matrix B as in (2.5.a) with $d_1 := b_{11} \mid c_{ij}$ for all $1 \leqslant i \leqslant m-1$ and $1 \leqslant j \leqslant n-1$. By induction, there exist $P' \in \mathrm{GL}_{m-1}(R)$ and $Q' \in \mathrm{GL}_{n-1}(R)$ such that

$$P'CQ' = \begin{pmatrix} d_2 & & & & \\ & \ddots & & & \\ & & d_k & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix} =: D'$$

with non-zero $d_2 \mid \cdots \mid d_k$. Therefore, with

$$P := \begin{pmatrix} 1 & 0 \\ 0 & P' \end{pmatrix} \quad \text{and} \quad Q := \begin{pmatrix} 1 & 0 \\ 0 & Q' \end{pmatrix}\,,$$

we obtain

$$PBQ = \begin{pmatrix} d_1 & & & & & \\ & d_2 & & & & \\ & & \ddots & & & \\ & & & d_k & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix}.$$

Now, by Remark 2.4 (b), any two gcds of the entries of C and D', respectively, are associates in R, so that d_1 also divides d_2. This finishes the first part of the proof.

Next we assume that two matrices

$$D = \begin{pmatrix} d_1 & & & & & \\ & d_2 & & & & \\ & & \ddots & & & \\ & & & d_k & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix} \text{ and } E = \begin{pmatrix} e_1 & & & & & \\ & e_2 & & & & \\ & & \ddots & & & \\ & & & e_l & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix}$$

in $M_{m\times n}(R)$ with non-zero $d_1, \ldots, d_k, e_1, \ldots, e_l$ satisfying $d_1 \mid \cdots \mid d_k$ and $e_1 \mid \cdots \mid e_l$ are equivalent. It will be convenient to set $d_{k+1} = d_{k+2} = \cdots = 0$ and $e_{l+1} = e_{l+2} = \cdots = 0$. Then $r_i = d_1 d_2 \cdots d_i$ and $s_i = e_1 e_2 \cdots e_i$ are gcds of the i-minors of D and E, respectively, and they are associates in R by Remark 2.4 (b). In particular, $r_i = 0$ if and only if $s_i = 0$. This implies $k = l$. Moreover, there exist units $u_1, \ldots, u_k$ of R with $s_i = r_i u_i$. Now it is easy to show by induction on i that $e_1 = d_1 u_1$ and $e_{i+1} = d_{i+1} u_{i+1} u_i^{-1}$ for $i \geqslant 2$. This concludes the proof of the Theorem. □

2.6 Corollary *Let R be a principal ideal domain and let M be a finitely generated R-module. Then there exist $s \in \mathbb{N}_0$, $k \in \mathbb{N}_0$, and non-units $d_1, \ldots, d_s \in R \smallsetminus \{0\}$ with $d_1 \mid \cdots \mid d_s$ such that*

$$M \cong R/d_1R \oplus \cdots \oplus R/d_sR \oplus R^{(k)}. \tag{2.6.a}$$

Proof Let M be generated by $y_1, \ldots, y_m \in M$ as R-module and let $f \in \mathrm{Hom}_R(R^{(m)}, M)$ be defined by $f(e_i) = y_i$. Then $U := \ker(f) \subseteq R^{(m)}$ is again R-free with $n := \mathrm{rk}_R(U) \leqslant m$ (by Theorem 2.2), and f is surjective, so $M \cong R^{(m)}/U$. Choose an ordered R-basis X of U and consider the matrix $A \in M_{m\times n}(R)$ representing the inclusion $U \subseteq R^{(m)}$ with respect to

X and $(e_1, \ldots, e_m)$. By Theorem 2.5 and since f is injective, there exist $d'_1, \ldots, d'_n \in R \setminus \{0\}$ with $d'_1 \mid \cdots \mid d'_n$ such that A is equivalent to the $(m \times n)$-matrix

$$\begin{pmatrix} d'_1 & & \\ & \ddots & \\ & & d'_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \end{pmatrix} .$$

This means that there exists an R-basis $f_1, \ldots, f_m$ of $R^{(m)}$ such that the elements $d'_1 f_1, \ldots, d'_n f_n$ form an R-basis of U. We set $d'_{n+1} = \cdots = d'_m := 0$ and obtain

$$M \cong R^{(m)}/U \cong (Rf_1 \oplus \cdots \oplus Rf_m)\big/(Rd'_1 f_1 \oplus \cdots \oplus Rd'_m f_m)$$

$$\cong R/d'_1 R \oplus \cdots \oplus R/d'_m R \cong R/d'_1 R \oplus \cdots \oplus R/d'_n R \oplus R^{(m-n)} .$$

Omitting those d'_is which are units, we obtain

$$M \cong R/d_1 R \oplus \cdots \oplus R/d_s R \oplus R^{(m-n)}$$

with $d_1 \mid \cdots \mid d_s$. □

2.7 Remark Let M be as in Corollary 2.6 and (2.6.a). Then $\mathrm{tor}_R(M) \cong R/d_1 R \oplus \cdots \oplus R/d_s R$. Moreover, there exists a free submodule U of M of R-rank k such that $M = \mathrm{tor}_R(M) \oplus U$.

2.8 Theorem (Chinese Remainder Theorem) *Let R be an arbitrary ring and let $I_1, \ldots, I_n$ be two-sided ideals of R such that $I_k + I_l = R$ for any $k \neq l$ from $\{1, \ldots, n\}$. Then the map*

$$\begin{aligned} R/I_1 \cap \ldots \cap I_n &\quad\to\quad R/I_1 \times \ldots \times R/I_n , \\ r + I_1 \cap \cdots \cap I_n &\quad\mapsto\quad (r + I_1, \ldots, r + I_n) , \end{aligned}$$

is a well-defined isomorphism of rings and R-modules.

Proof Consider the ring and R-module homomorphism

$$\rho \colon R \to R/I_1 \times \cdots \times R/I_n , \quad r \mapsto (r + I_1, \ldots, r + I_n) .$$

Its kernel is obviously $I_1 \cap \ldots \cap I_n$. Therefore, it suffices to show that ρ is surjective. Fix $k \in \{1, \ldots, n\}$. It suffices to find $b \in R$ such that $b + I_k = 1 + I_k$

and $b + I_l = 0 + I_l$ for all $l \neq k$. But for each $l \neq k$ there exist elements $a_l \in I_k$ and $b_l \in I_l$ such that $a_l + b_l = 1$. Therefore, $b := b_1 \cdots b_{k-1} b_{k+1} b_n \in I_l$ for all $l \neq k$. Moreover,

$$1 + I_k = (a_1 + b_1) \cdots (a_n + b_n) + I_k = b_1 \cdots b_n + I_k \,,$$

which completes the proof. □

2.9 Remark If R is a principal ideal domain and $d_1, \ldots, d_n \in R$ are pairwise coprime, then the ideals $I_1 := d_1 R, \ldots, I_n := d_n R$ satisfy the hypothesis of Theorem 2.8 and $I_1 \cap \ldots \cap I_n = d_1 \cdots d_n R$. Therefore, we obtain an isomorphism

$$\begin{aligned} \rho \colon R/d_1 \cdots d_n R &\to R/d_1 R \times \cdots \times R/d_n R \,, \\ r + d_1 \cdots d_n R &\mapsto (r + d_1 R, \ldots, r + d_n R) \,, \end{aligned}$$

of rings and R-modules.

2.10 Corollary *Let R be a principal ideal domain and let M be a finitely generated R-module. If $\mathcal{P}$ denotes a set of representatives for the primes of R up to associates, then*

$$M \cong R/p_1^{\alpha_1} R \oplus \cdots \oplus R/p_t^{\alpha_t} R \oplus R^{(l)} \tag{2.10.a}$$

for appropriate $t \in \mathbb{N}_0$, $l \in \mathbb{N}_0$, $p_1, \ldots, p_t \in \mathcal{P}$ and $\alpha_1, \ldots, \alpha_t \in \mathbb{N}$.

Proof This follows immediately from Corollary 2.6 and Remark 2.9, by decomposing each d_i into a product of powers of primes. □

2.11 Definition Let R be a principal ideal domain and let M be an R-module. For any prime p of R we set

$$M_p := \{ m \in M \mid p^\alpha m = 0 \text{ for some } \alpha \in \mathbb{N} \}$$

and call M_p the *p-component* or the *p-primary part* of M. We call M a *p-primary* module if $M = M_p$. Obviously, $f(M_p) \subseteq N_p$ for each map $f \in \mathrm{Hom}_R(M, N)$ to some R-module N. Note that $M_p \subseteq \mathrm{tor}_R(M)$ is an R-submodule and that $M_p = M_{p'}$ if p and p' are associates in R.

2.12 Theorem *Let R be a principal ideal domain, let M be an R-module, and let $\mathcal{P}$ be as in Corollary 2.10. Then*

$$\mathrm{tor}_R(M) = \bigoplus_{p \in \mathcal{P}} M_p \,.$$

Proof Obviously, $M_p \subseteq \mathrm{tor}_R(M)$. Conversely, let $0 \neq m \in \mathrm{tor}_R(M)$ and let $r \in R \smallsetminus \{0\}$ such that $rm = 0$. Then r is not a unit, and we can write $r = up_1^{\alpha_1} \cdots p_n^{\alpha_n}$ with a unit u of R, $p_1, \ldots, p_n \in \mathcal{P}$, and $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$. We may assume that $u = 1$ by choosing u^-1r instead of r if necessary. Set $r_i := r/p_i^{\alpha_i}$ for $i = 1, \ldots, n$. Then 1 is a gcd of $r_1, \ldots, r_n$ and there exist $a_1, \ldots, a_n \in R$ such that $a_1r_1 + \cdots + a_nr_n = 1$. Therefore, $m = a_1r_1m + \cdots + a_nr_nm$ with $a_ir_im \in M_{p_i}$ for each $i = 1, \ldots, n$. This implies that $\mathrm{tor}_R(M) \subseteq \sum_{p \in \mathcal{P}} M_p$.

Finally we show that the sum $\sum_{p \in \mathcal{P}} M_p$ is a direct sum. It suffices to show that if $m_1 + \cdots + m_n = 0$ with $m_i \in M_{p_i}$ for distinct $p_1, \ldots, p_n \in \mathcal{P}$, then $m_1 = \cdots = m_n = 0$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$ be such that $p_i^{\alpha_i}m_i = 0$ and define $r := p_1^{\alpha_1} \cdots p_n^{\alpha_n}$ and $r_i := r/p_i\alpha_i$ as above. Then $r_im_j = 0$ for all $i \neq j$. As above, there exist $a_1, \ldots, a_n \in R$ with $a_1r_1 + \cdots + a_nr_n = 1$. Therefore, $m_i = a_1r_1m_i + \cdots + a_nr_nm_i = a_ir_im_i = a_ir_i(m_1 + \cdots + m_n) = 0$. □

2.13 Lemma *Let R be a principal ideal domain and let p be a prime of R. Assume that $t, u \in \mathbb{N}_0$ and $1 \leqslant \alpha_1 \leqslant \cdots \leqslant \alpha_t$ and $1 \leqslant \beta_1 \leqslant \cdots \leqslant \beta_u$ are integers such that*

$$R/p^{\alpha_1}R \oplus \cdots \oplus R/p^{\alpha_t}R \cong R/p^{\beta_1}R \oplus \cdots \oplus R/p^{\beta_u}R \tag{2.13.a}$$

as R-modules. Then $t = u$ and $\alpha_i = \beta_i$ for all $i = 1, \ldots, t$.

Proof Let M be an R-module and let $k \in \mathbb{N}_0$. Then $p^{k+1}M \subseteq p^kM \subseteq M$ are R-submodules and $p^kM/p^{k+1}M$ is an R-module which is annihilated by pR. Therefore, $p^kM/p^{k+1}M$ is a vector space over the field $F := R/pR$ by setting $(r + pR)(m + p^{k+1}M) := rm + p^{k+1}M$ for $r \in R$ and $m \in p^kM$. If $M = R/p^\alpha R$ for some $\alpha \in \mathbb{N}$, we claim that

$$\dim_F(p^kM/p^{k+1}M) = \begin{cases} 0, \text{if } k \geqslant \alpha, \\ 1, \text{if } k < \alpha. \end{cases}$$

In fact, $p^kM = 0$ if $k \geqslant \alpha$, and $p^kM \cong p^kR/p^\alpha R$ if $k \leqslant \alpha$. Thus, for $k < \alpha$, we have

$$p^kM/p^{k+1}M \cong \big(p^kR/p^\alpha R\big)/\big(p^{k+1}R/p^\alpha R\big) \cong p^kR/p^{k+1}R \cong R/pR = F$$

as R-modules and therefore also as F-modules so that the claim is proven.

If we denote the left and right hand side of (2.13.a) by M and N, respectively, then for each $k \in \mathbb{N}_0$, the dimension $\dim_F(p^kM/p^{k+1}M)$ (resp. $\dim_F(p^kN/p^{k+1}N)$) counts the number of indices $i = 1, \ldots, t$ (resp. $j = 1, \ldots, u$) such that $\alpha_i > k$ (resp. $\beta_j > k$). But since $M \cong N$, this implies $t = u$ (by setting $k = 0$) and $\alpha_1 = \beta_1, \ldots, \alpha_t = \beta_t$ (by increasing k). □

2.14 Corollary *Let R be a principal ideal domain and let $\mathcal{D}$ be the set of products of elements in $\mathcal{P}$. Then each non-zero element of R which is not a unit has a unique associate in $\mathcal{D}$. If we write a finitely generated R-module M as*

$$M \cong R/d_1R \oplus \cdots \oplus R/d_sR \oplus R^{(k)}$$

as in Corollary 2.6 with non-units $d_1, \ldots, d_s \in R \smallsetminus \{0\}$ such that $d_1 \mid \cdots \mid d_s$, and as

$$M \cong R/p_1^{\alpha_1} \oplus \cdots \oplus R/p_t^{\alpha_t}R \oplus R^{(l)}$$

as in Corollary 2.10 with prime powers $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ of primes in $\mathcal{P}$, then $k = l$ and $k, d_1, \ldots, d_s$, $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ are uniquely determined (the prime powers up to permutations). The elements $d_1, \ldots, d_s$ are called the invariant factors *of M and $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ are called the* elementary divisors *of M.*

Proof We have

$$R/d_1R \oplus \cdots \oplus R/d_sR \cong \mathrm{tor}_R(M) \cong R/p_1^{\alpha_1} \oplus \cdots \oplus R/p_t^{\alpha_t},$$

and therefore,

$$R^{(k)} \cong M/\mathrm{tor}_R(M) \cong R^{(l)}$$

which implies $k = l$ and the uniqueness of k by Theorem 1.18. Moreover, by decomposing $\mathrm{tor}_R(M)$ into its p-primary parts (for $p \in \mathcal{P}$) we obtain the uniqueness of the prime powers $p_1^{\alpha_1}, \ldots, p_t^{\alpha_t}$ by Lemma 2.13, since

$$M_p \cong \bigoplus_{\substack{i \in \{1, \ldots, t\} \\ p_i = p}} R/p_i^{\alpha_i}R.$$

Moreover, for fixed $p \in \mathcal{P}$, the Chinese Remainder Theorem implies $(R/d_iR)_p \cong R/p^{\beta_i}R$, if d_i decomposes as $d_i = d_i'p^{\beta_i}$ and p does not divide d_i'. Therefore, $M_p \cong R/p^{\beta_1} \oplus \cdots \oplus R/p^{\beta_s}$ with $\beta_1 \leqslant \cdots \leqslant \beta_s$. Now Lemma 2.13 shows that $\beta_1, \ldots, \beta_s$ are uniquely determined by M. Therefore, $d_1, \ldots, d_s$ are uniquely determined. This completes the proof. □

2.15 Remark If $R = \mathbb{Z}$ one usually chooses the set of primes for $\mathcal{P}$. Then $\mathcal{D} = \{2, 3, 4, 5, 6, \ldots\}$. If $R = F[\lambda]$ for a field F one usually chooses the set of monic irreducible polynomials for $\mathcal{P}$. Then $\mathcal{D}$ is the set of monic polynomials of degree $\geqslant 1$.

2.16 Example (a) By Corollary 2.14, there are exactly two Abelian groups of order 20 (up to isomorphism), namely $\mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/5\mathbb{Z}$ (which is isomorphic to $\mathbb{Z}/20\mathbb{Z}$) and $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/5\mathbb{Z}$ (which is isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/10\mathbb{Z}$).

(b) Let

$$u_1 = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

and set $U := \langle u_1, u_2 \rangle_{\mathbb{Z}} \subseteq \mathbb{Z}^{(3)}$. Then $X := (u_1, u_2)$ is an ordered $\mathbb{Z}$-basis of U. We will try to find the invariant factors of $M := \mathbb{Z}^{(3)}/U$. We have to proceed as in the proof of Corollary 2.6. For this reason we have to start with the matrix A representing the inclusion $U \subseteq \mathbb{Z}^{(3)}$ with respect to some basis, say X and $E = (e_1, e_2, e_3)$. This gives

$$A = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ 2 & 3 \end{pmatrix}.$$

Next we have to transform A as in Theorem 2.5 into a diagonal matrix:

$$A = \begin{pmatrix} -2 & 1 \\ 0 & 2 \\ 2 & 3 \end{pmatrix} \xrightarrow{\bullet P_{12}} \begin{pmatrix} 1 & -2 \\ 2 & 0 \\ 3 & 2 \end{pmatrix} \xrightarrow{T_{21}(-2)\bullet} \begin{pmatrix} 1 & -2 \\ 0 & 4 \\ 3 & 2 \end{pmatrix} \xrightarrow{T_{31}(-3)\bullet}$$
$$\begin{pmatrix} 1 & -2 \\ 0 & 4 \\ 0 & 8 \end{pmatrix} \xrightarrow{\bullet T_{12}(2)} \begin{pmatrix} 1 & 0 \\ 0 & 4 \\ 0 & 8 \end{pmatrix} \xrightarrow{T_{32}(-2)\bullet} \begin{pmatrix} 1 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix} =: B.$$

This already shows that

$$M \cong \mathbb{Z}/1\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}$$

and that the invariant factors of M consist just of a single one, namely $d_1 = 4$. But we can derive more information from what we computed: B represents still the inclusion $U \subseteq \mathbb{Z}^{(3)}$ but with respect to different bases $Y = (v_1, v_2)$ and $E' = (e'_1, e'_2, e'_3)$ of U and $\mathbb{Z}^{(3)}$, respectively. In order to compute Y and E' we have to write B in the form $B = T^{-1}AS$, and then have

$$\begin{gathered}(v_1, v_2) = (u_1, u_2)S = (u_1, u_2)P_{12}T_{12}(2) = \\ (u_2, u_1)T_{12}(2) = (u_2, u_1 + 2u_2).\end{gathered}$$

and

$$\begin{gathered}(e'_1, e'_2, e'_3) = (e_1, e_2, e_3)T = (e_1, e_2, e_3)T_{21}(2)T_{31}(3)T_{32}(2) = \\ (e_1 + 2e_2, e_2, e_3)T_{31}(3)T_{32}(2) = (e_1 + 2e_2 + 3e_3, e_2, e_3)T_{32}(2) = \\ (e_1 + 2e_2 + 3e_3, e_2 + 2e_3, e_3).\end{gathered}$$

This also shows that in the above isomorphism $M \cong \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}$ the element in M corresponding to $(1 + 4\mathbb{Z}, 0)$ is $e'_2 + U$ and the one corresponding to $(0 + 4\mathbb{Z}, 1)$ is $e'_3 + U$.

3 Normal Forms of Matrices

3.1 Notation Throughout this section we fix a field F, a finite dimensional F-vector space V, an ordered F-basis $X = (v_1, \dots, v_n)$ of V, and an endomorphism $f \in \mathrm{End}_F(V)$. We set $A := {}_X M_X(f) = (a_{ij}) \in M_n(F)$, i.e.,

$$f(v_i) = \sum_{j=1}^{n} a_{ji} v_j \,, \qquad i = 1, \dots, n \,.$$

We put $R := F[\lambda]$ and view V a an R-module via $g(\lambda) \cdot v := (g(f))(v)$ for $g(\lambda) \in R$ and $v \in V$.

3.2 Remark (a) If $g(\lambda) = a_m \lambda^m + \cdots + a_1 \lambda + a_0 \in R$ with $a_m \neq 0$, then $R/g(\lambda)R$ is an F-vector space with F-basis

$$1 + g(\lambda)R,\ \lambda + g(\lambda)R,\ \dots,\ \lambda^{m-1} + g(\lambda)R \,.$$

In fact, dividing an arbitrary $h(\lambda) \in R$ by $g(\lambda)$ gives $h(\lambda) = q(\lambda)g(\lambda) + r(\lambda)$ with $q(\lambda), r(\lambda) \in R$ and $\deg(r(\lambda)) < m$, so that $h(\lambda) + g(\lambda)R = r(\lambda) + g(\lambda)R$. Moreover, if $b_0(1 + g(\lambda)R) + \cdots + b_{m-1}(\lambda^{m-1} + g(\lambda)R) = 0$, then $b_0 + b_1\lambda + \cdots b_{m-1}\lambda^{m-1} \in g(\lambda)R$ which implies $b_0 = \cdots = b_{m-1} = 0$, since $\deg(g(\lambda)) = m$.

(b) If $g(\lambda) = (\lambda - r)^m$ for some $r \in F$, then also

$$1 + g(\lambda)R,\ (\lambda - r) + g(\lambda)R,\ \dots,\ (\lambda - r)^{m-1} + g(\lambda)R$$

form an F-basis of $R/g(\lambda)R$, since these m elements of $R/g(\lambda)R$ are F-linear independent by a similar argument as in (a).

(c) V is a finitely generated R-torsion module, since it is finite dimensional over F. Suppose that $V \cong R/p_1(\lambda)^{\alpha_1}R \oplus \cdots \oplus R/p_t(\lambda)^{\alpha_t}R$ with monic irreducible $p_i(\lambda) \in R$ and $\alpha_i \in \mathbb{N}$. Then $n = \dim_F(V) = n_1 + \cdots + n_t$ with $n_i = \alpha_i \deg(p_i(\lambda))$, and V (as R-module) has generators $z_1, \dots, z_t \in V$ corresponding to the elements $(0, \dots, 0, 1 + p_i(\lambda)^{\alpha_i}R, 0, \dots, 0)$ for $i = 1, \dots, t$. Then, by (a),

$$Z = (z_1, f(z_1), \dots, f^{n_1 - 1}(z_1), z_2, f(z_2), \dots)$$

is an F-basis of V and we have

$${}_Z M_Z(f) = \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_t \end{pmatrix}$$

with

$$B_i = \begin{pmatrix} 0 & \cdots & \cdots & 0 & -a_0^{(i)} \\ 1 & \ddots & & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n_i - 1}^{(i)} \end{pmatrix} \in M_{n_i}(F) \,, \quad i = 1, \dots, t,$$

if $p_i(\lambda)^{\alpha_i} = \lambda^{n_i} + a^{(i)}_{n_i-1} + \cdots + a^{(i)}_0$. Any matrix like the B_is, such that the polynomial formed from the last column is a power of an irreducible polynomial of R is called an *F-rational block*. Any block diagonal matrix with F-rational blocks is called *F-rational*. The matrix ${}_ZM_Z(f)$ is called the *rational normal form* of f or A. One can show that up to reordering the blocks B_i it is uniquely determined by f or A. That means, if $S^{-1}AS$ is again F-rational for some $S \in \mathrm{GL}_n(F)$, then the blocks of $S^{-1}AS$ coincide with the $B_1, \ldots, B_t$ above.

(d) If $p_i(\lambda)^{\alpha_i} = (\lambda - r_i)^{n_i}$ for each $i = 1, \ldots, t$ and some $r_1, \ldots, r_t \in F$ (note that this always is the case for $F = \mathbb{C}$), and if $z_1, \ldots, z_t \in V$ are as in (c), then we choose the F-basis

$$\begin{aligned} Y &= (z_1,\ (\lambda - r_1)z_1,\ (\lambda - r_1)^2 z_1,\ \ldots,\ (\lambda - r_1)^{n_1-1} z_1,\ z_2,\ (\lambda - r_2)z_2,\ \ldots) \\ &= (z_1,\ (f - r_1\mathrm{id}_V)(z_1),\ (f - r_1\mathrm{id}_V)(z_1),\ \ldots) \end{aligned}$$

of V and obtain

$${}_YM_Y(f) = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_t \end{pmatrix}$$

with

$$J_i = \begin{pmatrix} r_i & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & r_i \end{pmatrix} \in M_{n_i}(F)\,, \quad i = 1, \ldots, t.$$

Matrices of the form J_i are called *Jordan blocks*, matrices of the form ${}_YM_Y(f)$ are called *Jordan matrices* and ${}_YM_Y(f)$ is called the *Jordan normal form* of f or A. Again, the Jordan normal form of A or f is unique up to permutation of the blocks in the same sense as the rational normal form. Note that the rational and Jordan (if existent) normal forms are already determined by the R-module structure of V.

3.3 Theorem *The invariant factors of the R-module V are the monic non-unit invariant factors of $\lambda \cdot 1_n - A \in M_n(R)$.*

Proof According to the proof of Corollary 2.6, the invariant factors of V are the non-unit invariant factors of the matrix representing the inclusion $\ker(\rho) \subseteq R^{(m)}$, where $\rho\colon R^{(m)} \to V$ is any surjective R-linear map. We choose

$$\rho\colon R^{(n)} \to V\,, \quad e_i \mapsto v_i\,,$$

and claim that the elements

$$f_i := \lambda e_i - \sum_{j=1}^{n} a_{ji} e_j \in R^{(n)}\,, \quad i = 1, \ldots, n\,,$$

form an R-basis of $\ker(\rho)$. Then, the matrix representing the inclusion $\ker(\rho) \subseteq R^{(n)}$ with respect to $(f_1, \dots, f_n)$ and $(e_1, \dots, e_n)$ is just $\lambda \cdot 1_n - A$, and the proof would be complete.

Since $\rho(f_i) = f(v_i) - \sum_{j=1}^n a_{ji} v_j = 0$, we have $f_1, \dots, f_n \in \ker(\rho)$. Since $\lambda e_i = f_i + \sum_{j=1}^n a_{ji} e_j$, we have $Re_i \subseteq Rf_i + \sum_{j=1}^n Fe_j$, and so $R^{(n)} = \sum_{i=1}^n Rf_i + \sum_{i=1}^n Fe_i$. If $u \in \ker(\rho)$ and $u = \sum_{i=1}^n h_i(\lambda) f_i + \sum_{i=1}^n b_i e_i$, then applying ρ to this equation gives $0 = \rho(\sum_{i=1}^n b_i e_i) = \sum_{i=1}^n b_i v_i$. But this implies $b_1 = \dots = b_n = 0$, and so $u \in \langle f_1, \dots, f_n \rangle_R$. Moreover, if $\sum_{i=1}^n h_i(\lambda) f_i = 0$ with $h_i(\lambda) \in R$, then

$$\sum_{i=1}^n \lambda h_i(\lambda) e_i = \sum_{i=1}^n h_i(\lambda) \sum_{j=1}^n a_{ji} e_j \,,$$

and so

$$\lambda h_k(\lambda) = \sum_{i=1}^n h_i(\lambda) a_{ki} \,,$$

for each $k = 1, \dots, n$. If one of the $h_k(\lambda)$s is non-zero and if we choose $l \in \{1, \dots, n\}$ such that $h_l(\lambda)$ has maximal degree, then we obtain $\lambda h_l(\lambda) = \sum_{i=1}^n h_i(\lambda) a_{li}$ which is impossible, since the two sides have different degrees. $\square$

3.4 Remark In order to find an F-basis $W = (w_1, \dots, w_n)$ of V such that ${}_W M_W(f)$ is a Jordan matrix or an F-rational matrix, we have to determine T in

$$T^{-1}(\lambda 1_n - A) S = D = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & d_1(\lambda) & & \\ & & & & \ddots & \\ & & & & & d_s\lambda \end{pmatrix}$$

with non-constant polynomials $d_1(\lambda) \mid \cdots \mid d_s(\lambda)$ in R, and with $S, T \in \mathrm{GL}_n(R)$. Then the elements $u_1, \dots u_s \in V$ which are determined by

$$(v_1, \dots, v_n) T = (*, \dots, *, u_1, \dots, u_s)$$

generate V as R-module such that $V = Ru_1 \oplus \dots \oplus Ru_s$ with $\mathrm{ann}_R(u_i) = d_i(\lambda) R$, so that $Ru_i \cong R/d_i(\lambda)R$ under $u_i \leftrightarrow 1 + d_i(\lambda)R$. Going through the proof of the Chinese Remainder Theorem for each $R/d_i(\lambda)R$ then gives the elements $z_1, \dots, z_t \in V$ $(t \geqslant s)$ as in Remark 3.2 (c), and we can continue as in Remark 3.2.

3.5 Remark (a) Let $d_1(\lambda) \mid \cdots \mid d_s(\lambda)$ be the invariant factors of $\lambda 1_n - A \in M_n(R)$. Then $d_s(\lambda)$ is the minimal polynomial $m_A(\lambda)$ of A (since $g(A) = 0 \iff g(f)v_i = 0$ for all $i = 1, \ldots, n \iff g(\lambda)V = 0 \iff g(\lambda) \in \mathrm{ann}_R(V) = d_s(\lambda)R \iff d_s(\lambda) \mid g(\lambda)$). Moreover, the product $d_1(\lambda) \cdots d_s(\lambda)$ is the characteristic polynomial $\mathrm{ch}_A(\lambda)$ of A (since $\det(\mathrm{diag}(1, \ldots, 1, d_1(\lambda), \ldots, d_s(\lambda)))$ and $\det(\lambda 1_n - A)$ are associates by Theorem 3.3 and both are monic). In particular this shows that $m_A(\lambda)$ divides $\mathrm{ch}_A(\lambda)$ and that they have the same irreducible factors.

(b) We call two matrices $A, B \in M_n(F)$ *similar* if there exists $S \in \mathrm{GL}_n(F)$ such that $B = S^{-1}AS$. It is not difficult to show that the following are equivalent:

(i) A and B are similar.

(ii) A and B have the same rational normal form (up to reordering the blocks).

(iii) $\lambda 1_n - A$ and $\lambda 1_n - B$ have the same monic invariant factors.

If $F = \mathbb{C}$, the rational normal form in (ii) can be replaced by the Jordan normal form.

II. Field Extensions and Galois Theory

1 Questions

We will address the following questions in the course of this chapter:

- Are there formulae giving the solutions of a polynomial equation $X^n + a_{n-1}X^{n-1} + \cdots a_1 X + a_0$ (with $a_0, \ldots, a_{n-1}$ from a field) in terms of the operations $+, -, \cdot, /$, and $\sqrt[k]{}$. Note that we can do this in the case $n = 2$: The solutions of $X^2 + aX + b$ are
$$\frac{1}{2}(-a \pm \sqrt{a^2 - 4b}).$$
- Let $F \subseteq E$ be fields. Are there finitely or infinitely many intermediate fields between E and F? How can we determine them?
- Geometric problems: Suppose we have only a straight edge and a compass.
  - Can we square the circle?
  - Can we trisect any given angle?
  - Can we construct a regular n-gon?
  - Can we double a cube?
  - What numbers in the complex plane can we construct, starting from 0 and 1?

2 Basic Definitions and Properties

2.1 Definition Let $F \subseteq E$ be fields (we also say that E/F is a *field extension*) and let $x \in E$. If there exists $0 \neq f \in F[X]$ such that $f(x) = 0$, we call x *algebraic over* F; otherwise we call x *transcendental over* F. If x is algebraic over F, the polynomials $g \in F[X]$ with $g(x) = 0$ form a non-zero ideal of $F[X]$. The monic generator $\min_{x/F} \in F[X]$ of this ideal is called the *minimal polynomial of* x *over* F. It has the property that $\min_{x/F}(x) = 0$ and for every $g \in F[X]$ with $g(x) = 0$ one has $\min_{x/F} \mid g$. Obviously, $\min_{x/F}$ is irreducible in $F[X]$.

2.2 Example The element $\sqrt{2}$ is algebraic over $\mathbb{Q}$ with minimal polynomial $\min_{\sqrt{2}/\mathbb{Q}} = X^2 - 2$. The minimal polynomial of 5 over $\mathbb{Q}$ is $X - 5$.

2.3 Remark Let E/F be a field extension.

(a) Recall that if there exists an $n \in \mathbb{N}$ such that $n \cdot 1_F = 1 + \cdots + 1 = 0$, then the minimal such n is a prime p, called the *characteristic* of F and denoted by $\mathrm{char}(F)$. If no such n exists, then we set $\mathrm{char}(F) = 0$.

(b) Note that E is an F-vector space. We set $[E : F] := \dim_F E$ and call $[E : F]$ the *degree* of the extension E/F. It is easy to see that if E'/E is a further field extension, if $\{x_i \mid i \in I\}$ is an F-basis of E, and if $\{y_j \mid j \in J\}$ is an E-basis of E', then $\{x_i y_j \mid (i,j) \in I \times J\}$ is an F-basis of E'. In particular, E'/F is of finite degree if and only if E'/E and E/F are of finite degree. In this case we have $[E' : F] = [E' : E][E : F]$.

(c) For $x \in E$ we write $F[x]$ for the smallest subring of E containing F and x. It is easy to see that this is the set

$$\{f(x) \mid f \in F[X]\}.$$

Moreover, we write $F(x)$ for the smallest subfield of E containing F and x. It is easy to see that this is the set

$$\{f(x)/g(x) \mid f, g \in F[X], g(x) \neq 0\}.$$

More generally, for any subset $\{x_i \mid i \in I\}$ of E we define $F[x_i \mid i \in I]$ (resp. $F(x_i \mid i \in I)$) to be the smallest subring (resp. subfield) of E containing F and $\{x_i \mid i \in I\}$. These are given by

$$\{f(x_{i_1}, \ldots, x_{i_n}) \mid n \in \mathbb{N},\ i_1, \ldots, i_n \in I,\ f \in F[X_1, \ldots, X_n]\}$$

and

$$\{f(x_{i_1}, \ldots, x_{i_n})/g(x_{j_1}, \ldots, x_{j_m}) \mid m, n \in \mathbb{N},\ i_1, \ldots, i_n, j_1, \ldots, j_m \in I,\\ f \in F[X_1, \ldots, X_n],\ g \in F[X_1, \ldots, X_m],\ g(x_{j_1}, \ldots, x_{j_m}) \neq 0\}.$$

(d) Note that a field F has only the ideals 0 and F. Therefore, each ring homomorphism $f\colon F \to R$ to any ring R is injective ($f(1) = 1$).

2.4 Theorem *Let E/F be a field extension and let $x \in E$ be algebraic with minimal polynomial $m(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0$.*

(a) The map $\rho\colon F[X] \to F[x]$, $f \mapsto f(x)$, is a surjective ring homomorphism with kernel $(m(X))$.

(b) The ring $F[x]$ is a field; in particular, $F[x] = F(x)$.

(c) We have $[F(x) : F] = n$, and $1, x, x^2, \ldots, x^{n-1}$ is an F-basis of $F(x)$.

Proof (a) It is easy to check that ρ is a ring homomorphism. It is surjective by the description of $F[x]$ in Remark 2.3 (c).

(b) By (a) we have a ring isomorphism $\bar{\rho}\colon F[X]/(m(X)) \to F[x]$. Since $m(X)$ is irreducible, both rings are fields.

(c) We know that $1+(m(X)), X+(m(X)), \ldots, X^{n-1}+(m(X))$ is an F-basis of $F[X]/(m(X))$. Since $\bar{\rho}$ is also F-linear, $\bar{\rho}(X^i+(m(X))) = x^i$, $i = 0, \ldots, n-1$, form an F-basis of $F(x)$. □

2.5 Definition A field extension E/F is called *algebraic*, if every element of E is algebraic over F.

2.6 Theorem *Let E/F be a field extension.*

(a) For $x \in E$ we have: x is algebraic over F $\iff$ $[F(x) : F] < \infty$.

(b) If $[E : F] < \infty$, then E is algebraic over F.

(c) If $E = F(x_i \mid i \in I)$ with elements $x_i \in E$, $i \in I$, which are algebraic over F, then E is algebraic over F.

Proof (a) $\Rightarrow$: This follows from Theorem 2.4. $\Leftarrow$: If $[F(x) : F] = n$, then the set $1, x, x^2, \ldots, x^n$ is F-linearly dependent and we obtain a relation $a_nx^n + \cdots a_1x + a_0 = 0$ with $a_i \in F$, not all of them being zero.

(b) This is clear with (a).

(c) Let $y \in E = F(x_i \mid i \in I)$. Then $y \in F(y_1, \ldots, y_k)$ for some finite subset $\{y_1, \ldots, y_k\}$ of $\{x_i \mid i \in I\}$. Since y_i is algebraic over F it is also algebraic over $F(y_1, \ldots, y_{i-1})$, $i = 1, \ldots, k$. Therefore, $[F(y_1, \ldots, y_i) : F(y_1, \ldots, y_{i-1})]$ is finite for all $i = 1, \ldots, k$ by (a). Now Remark 2.3 (b) implies that $[F(y_1, \ldots, y_k) : F]$ is finite. Hence, y is algebraic over F by (b). □

3 The Field of Constructible Elements

3.1 Definition Let $z_1, \ldots, z_k \in \mathbb{C}$. We set $C_0(z_1, \ldots, z_k) := \{0, 1, z_1, \ldots, z_k\}$ and define $C_{n+1}(z_1, \ldots, z_k) \subseteq \mathbb{C}$, for $n \in \mathbb{N}_0$, recursively as the union of $C_n(z_1, \ldots, z_k)$ and the complex numbers arising as intersections of

(a) two different lines, each one drawn through two elements of $C_n(z_1, \ldots, z_k)$, or

(b) two different circles, each one having an element of $C_n(z_1, \ldots, z_k)$ as center and the distance between two elements of $C_n(z_1, \ldots, z_k)$ as radius, or

(c) a line and a circle as in (a) and (b).

We call $C(z_1, \dots, z_k) := \bigcup_{n\in\mathbb{N}_0} C_n(z_1, \dots, z_k)$ the set of elements *constructible* from $\{z_1, \dots, z_k\}$. We will write just C_n and C if $k = 0$ (i.e., if we start with $C_0 = \{0, 1\}$).

3.2 Theorem *Let $z_1, \dots, z_k \in \mathbb{C}$. The set $C(z_1, \dots, z_k)$ is the smallest subfield E of $\mathbb{C}$ which contains $z_1, \dots, z_k$ and is closed under complex conjugation ($z \in E \Rightarrow \bar{z} \in \mathbb{C}$) and closed under taking square roots ($z \in \mathbb{C}$, $z^2 \in E \Rightarrow z \in E$).*

Proof (a) $C(z_1, \dots, z_k)$ is a field: Let $x, y \in C(z_1, \dots, z_k)$. It is not difficult to show that $x + y$, $-x$, $x \cdot y$, and x^{-1} (if $x \neq 0$) arise after finitely many of the steps (a)-(c) of Definition 3.1. For details see [J, pp. 217–218].

(b) $C(z_1, \dots, z_k)$ is closed under complex conjugation, since we can construct the reflection of a given point at the real axis.

(c) $C(z_1, \dots, z_k)$ is closed under taking square roots: Suppose $z \in \mathbb{C}$ with $z^2 = x \in C(z_1, \dots, z_k)$. Then z arises from bisecting one of the two angles between the positive real ray and the ray with origin in 0 through x, and by taking the square root of the length of x. Both can be constructed.

(d) Let $E \subseteq \mathbb{C}$ be a subfield which contains $z_1, \dots, z_k$ and is closed under complex conjugation and under taking square roots. Then $i \in E$, since $i^2 = -1 \in E$. Moreover, for $a = \alpha_1 + \alpha_2 i \in \mathbb{C}$ with $\alpha_1, \alpha_2 \in \mathbb{R}$, we have: $a \in E \iff \alpha_1 \in E$ and $\alpha_2 \in E$. In fact $\alpha_1 = (a + \bar{a})/2$ and $\alpha_2 = (a - \bar{a})/2i$. We will show by induction on $n \in \mathbb{N}_0$ that $C_n(z_1, \dots, z_k) \subseteq E$. Obviously, $C_0(z_1, \dots, z_k) = \{0, 1, z_1, \dots, z_k\} \subseteq E$. Assume that $C_n(z_1, \dots, z_k) \subseteq E$. Any line through elements of $C_n(z_1, \dots, z_k)$ is given as the set of solutions $x = \xi_1 + \xi_2 i$ of the equation $\xi_2 = \alpha\xi_1 + \beta$ or as the set of solutions of the equation $\xi_1 = \gamma$ for some $\alpha, \beta, \gamma \in E \cap \mathbb{R}$. Note that the distance ρ between any two elements of $C_n(z_1, \dots, z_k)$ is in E. Therefore, an arbitrary circle with center at $m = \mu_1 + \mu_2 i \in C_n(z_1, \dots, z_k)$ and radius ρ which is the distance between two points of $C_n(z_1, \dots, z_k)$ is the set of solutions $x = \xi_1 + \xi_2 i$ of the equation $(\xi_1 - \mu_1)^2 + (\xi_2 - \mu_2)^2 = \rho^2$. Therefore, the intersections of any two such lines or circles are solutions of linear or quadratic equations with coefficients from E. Therefore, its solutions are still elements in E and we have $C_{n+1}(z_1, \dots, z_k) \subseteq E$. □

3.3 Definition Let E/F be a field extension. A *square root tower* in E over F is a finite series of field extensions

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_l \subseteq E$$

such that there exist elements $u_i \in F_i$ with $F_i = F_{i-1}(u_i)$ and $u_i^2 \in F_{i-1}$ for all $i = 1, \dots, l$. In particular, $F_i = F(u_1, \dots, u_i)$ and $[F_i : F] \mid 2^i$ for all $i = 1, \dots, l$.

3.4 Theorem *Let $z_1, \ldots, z_k \in \mathbb{C}$ and set $F := \mathbb{Q}(z_1, \ldots, z_k, \bar{z}_1, \ldots, \bar{z}_k)$. For $z \in \mathbb{C}$ the following are equivalent:*

(i) *$z \in C(z_1, \ldots, z_k)$.*

(ii) *There exists a square root tower $F = F_0 \subseteq \ldots \subseteq F_l$ in $\mathbb{C}$ over F such that $z \in F_l$.*

Proof (ii) $\Rightarrow$ (i): Let $F = F_0 \subseteq \ldots \subseteq F_l$ be a square root tower as in (ii). Then F_l is contained in every subfield E of $\mathbb{C}$ which contains $z_1, \ldots, z_k$ and which is closed under complex conjugation and under taking square roots. Hence, by Theorem 3.2, F_l is contained in $C(z_1, \ldots, z_k)$.

(i) $\Rightarrow$ (ii): Let E denote the set of elements $z \in \mathbb{C}$ which satisfy (ii). By Theorem 3.2, it suffices to show that E is a field which contains $z_1, \ldots, z_k$ and which is closed under complex conjugation and under taking square roots. Obviously, E contains $z_1, \ldots, z_k$. Moreover, if $x, y \in E$ and if $x \in F_l$, $y \in E_m$, with square root towers $F = F_0 \subseteq \cdots \subseteq F_l$ and $F = E_0 \subseteq \cdots \subseteq E_m$ such that $F_i = F_{i-1}(u_i)$, $E_j = E_{j-1}(v_j)$ for elements $u_i, v_j \in E$ satisfying $u_i^2 \in F_{i-1}$ and $v_j^2 \in E_{j-1}$ for all $1 \leqslant i \leqslant l$ and $1 \leqslant j \leqslant m$, then $F = F_0 \subseteq \cdots \subseteq F_l \subseteq F_l(v_1) \subseteq F_l(v_1, v_2) \subseteq \cdots \subseteq F_l(v_1, \ldots, v_m)$ is a square root tower containing $x+y$, xy, $-x$, and x^{-1} (if $x \neq 0$). This shows that E is a field. E is also closed under taking square roots, since we just have to extend a square root tower by one step. To prove that E is closed under taking complex conjugation, assume that $z \in F_l$ and that $F = F_0 \subseteq \cdots \subseteq F_l \subseteq \mathbb{C}$ is a square root tower with $F_i = F_{i-1}(u_i)$ and $u_i^2 \in F_{i-1}$ for all $i = 1, \ldots, l$. Then, $\bar{z} \in \bar{F}_l$ and $F = \bar{F} = \bar{F}_0 \subseteq \cdots \subseteq \bar{F}_l \subseteq \mathbb{C}$ is a square root tower with $\bar{F}_i = \bar{F}_{i-1}(\bar{u}_i)$ and $\bar{u}_i^2 \in \bar{F}_{i-1}$ for all $i = 1, \ldots, l$. $\square$

3.5 Corollary *Let $z_1, \ldots, z_k \in \mathbb{C}$ and assume that $z \in C(z_1, \ldots, z_k)$. Then $[F(z) : F]$ is a power of 2, where $F = \mathbb{Q}(z_1, \ldots, z_k, \bar{z}_1, \ldots, \bar{z}_k)$.*

Proof This follows immediately from Theorem 3.4. $\square$

3.6 Theorem *The duplication of the cube is impossible.*

Proof Assume that it is possible. Then $\sqrt[3]{2} \in C$ and $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = \deg(\min_{\sqrt[3]{2}/\mathbb{Q}}(X))$ is be a power of 2 by Corollary 3.5. But $\min_{\sqrt[3]{2}/\mathbb{Q}}(X) = X^3 - 2$ since it is irreducible in $\mathbb{Z}[X]$ (by Eisenstein's criterion), and then also in $\mathbb{Q}[X]$ (cf. [J, 2.16, Lemma 3]). This is a contradiction. $\square$

3.7 Theorem *It is not possible to trisect an arbitrary angle.*

Proof Assume that this is possible. Then $\zeta_{18} := \cos 20° + i \sin 20° \in C$, since $\zeta_6 = \cos 60° + i \sin 60° \in C$. But with ζ_{18} also $a := \cos 20° = (\zeta_{18} - \bar{\zeta}_{18})/2 \in C$. Since $\cos(3\theta) = 4\cos^3\theta - 3\cos\theta$, we have $4a^3 - 3a = \cos 60° = 1/2$. With a also $b := 2a - 1$ is in C and we have

$$0 = 4\Big(\frac{b+1}{2}\Big)^3 - 3\Big(\frac{b+1}{2}\Big) - \frac{1}{2} = \frac{1}{2}(b^3 + 3b^2 - 3)\,,$$

and so $b^3+3b^2-3=0$. But X^3+3X^2-3 is irreducible in $\mathbb{Z}[X]$ by Eisenstein's criterion, and then also in $\mathbb{Q}[X]$ (cf. [J, 2.16, Lemma 3]). Therefore, this is the minimal polynomial of $b \in C$ over $\mathbb{Q}$ and $[\mathbb{Q}(b):\mathbb{Q}]=3$. This contradicts Corollary 3.5. $\square$

4 Splitting Fields and Normal Field Extensions

4.1 Definition Let F be a field. An F-algebra (R,i) is a ring R together with a ring homomorphism $i\colon F \to R$ such that $i(F) \subseteq Z(R)$. An F-algebra homomorphism between two F-algebras (R,i) and (S,j) is a ring homomorphism $\varphi\colon R \to S$ such that $f(i(\alpha)) = j(\alpha)$ for all $\alpha \in F$. Note that R carries the structure of an F-vector space via $\alpha \cdot r := i(\alpha)r$ for $\alpha \in F$ and $r \in R$. Often we identify F via i with its image $i(F) = F \cdot 1_R$ in R. Then we have $\varphi(\alpha) = \alpha$ for any F-algebra homomorphism φ and $\alpha \in F$. Often we only write R instead of (R,i) if the map i is obvious from the context. We denote the set of F-algebra homomorphisms between R and S by $\mathrm{Hom}_{F-\mathsf{alg}}(R,S)$. If (E,i) is an F-algebra and also a field, then we may consider E/F as a field extension by identifying F with $i(F)$.

4.2 Examples (a) If E/F is a field extension, then E is an F-algebra.

(b) The polynomial ring $F[X_i \mid i \in I]$ in the indeterminate X_i, $i \in I$, is an F-algebra.

(c) The matrix ring $M_n(F)$ is an F-algebra with $i(\alpha) = \alpha \cdot 1_n$ for $\alpha \in F$.

4.3 Theorem *Let F be a field and let X_i, $i \in I$, be a family of indeterminates. If R is a commutative F-algebra and r_i, $i \in I$, is a family of elements in R, then there exists a unique $\varphi\colon \mathrm{Hom}_{F-\mathsf{alg}}(F[X_i \mid i \in I], R)$ such that $\varphi(X_i) = r_i$.*

Proof The uniqueness part follows from the fact that each element of $F[X_i \mid i \in I]$ is an F-linear combination of products of the elements X_i, $i \in I$. In order to show the existence of φ, let K be the set of all functions $k\colon I \to \mathbb{N}_0$ such that $k(i) \neq 0$ for only finitely many $i \in I$. Then the elements $X^k := \prod_{i\in I} X_i^{k(i)}$, $k \in K$ form an F-basis of $F[X_i \mid i \in I]$. Therefore, there exists an F-linear map $\varphi\colon F[X_i \mid i \in I] \to R$ such that $\varphi(X^k) = \prod_{i \in I} r_i^{k(i)}$, for all $k \in K$. Now it is easy to see that

$$\varphi(X^k X^l) = \prod_{i\in I} r_i^{k(i)+l(i)} = \varphi(X^k)\varphi(X^l)$$

for all $k, l \in K$. Note that $\varphi(1) = 1$. For arbitrary elements $\sum_{p=1}^{n} \alpha_p X^{k_p}$ and $\sum_{q=1}^{m} \beta_q X^{l_q}$ of $F[X_i \mid i \in I]$ with $k_p, l_q \in K$ and $\alpha_p, \beta_q \in F$ for $1 \leqslant p \leqslant n$ and

$1 \leqslant q \leqslant m$, we have:

$$\varphi((\sum_{p=1}^{n} \alpha_p X^{k_p})(\sum_{q=1}^{m} \beta_q X^{l_q})) = \sum_{p=1}^{n}\sum_{q=1}^{m} \alpha_p \beta_q \varphi(X^{k_p} X^{l_q})$$
$$= \sum_{p=1}^{n}\sum_{q=1}^{m} \alpha_p \beta_q \varphi(X^{k_p})(\varphi(X^{l_q}) = \varphi(\sum_{p=1}^{n} \alpha_p X^{k_p})\varphi(\sum_{q=1}^{m} \beta_q X^{l_q}) ,$$

which shows that φ is a ring homomorphism. Moreover, $\varphi(\alpha) = \varphi(\alpha \cdot 1) = \alpha\varphi(1) = \alpha 1_R$, for each $\alpha \in F$. This implies that $\varphi \in \mathrm{Hom}_{F-\mathsf{alg}}(F[X_i \mid i \in I], R)$. □

4.4 Definition Let F be a field and let $f \in F[X]$ with $\deg(f) = n > 0$. An extension field E of F is called a *splitting field of f over F* if f decomposes as $f(X) = \alpha(X - r_1)\cdots(X_r - n)$ in $E[X]$ (with $\alpha \in F$ and $r_1, \ldots, r_n \in E$) and if $E = F(r_1, \ldots, r_n)$. Note that then E/F is a *finite* (i.e., $[E : F] < \infty$) and algebraic field extension.

4.5 Remark Let F be a field and let $f \in F[X]$ with $\deg(f) > 0$. Then there exists an extension field E of F of finite degree such that f has a root $r \in E$. In fact, let $p(X) = \alpha_n X^n + \ldots + \alpha_1 X + \alpha_0$ be an irreducible factor of f in $F[X]$. Then $E := F[X]/(p)$ can be considered as an extension field of F and for $r := X + (p) \in E$ we have:

$$p(r) = \alpha_n(X + (p))^n + \ldots + \alpha_1(X + (p)) + \alpha_0$$
$$= \alpha_n X^n + \ldots + \alpha_0 + (p) = p(X) + (p) = 0 \in E\,.$$

Therefore, $f(X) = g(X)(X - r)$ with some $g \in E[X]$.

4.6 Theorem *Let F be a field and let $f \in F[X]$ with $\deg(f) > 0$. Then there exists a splitting field E of f over F.*

Proof Induction on $n = \deg(f)$. If $n = 1$ we may take $E = F$. So let $\deg(f) \geqslant 2$. By Remark 4.5, there exists a finite extension field E_1 of F such that $f(X) = (X - r_1)g(X)$ with $g \in E_1[X]$ and $r_1 \in E_1$. Since $\deg(g) < n$, by induction there exists a finite field extension E_2/E_1 such that $g(X) = \alpha(X - r_2)\cdots(X - r_{n-1})$ with $r_2, \ldots, r_n \in E_2$ and $\alpha \in F$. Therefore, $f(X) = \alpha(X - r_1)\cdots(X - r_n)$ in $E_2(X)$ and $E := F(r_1, \ldots, r_n)$ is a splitting field of f over F. □

4.7 Example A splitting field of $f(X) = X^3 - 2$ over $\mathbb{Q}$ is $E := \mathbb{Q}(\sqrt[3]{2}, \zeta_3\sqrt[3]{2}, \zeta_3^2\sqrt[3]{2})$ with $\zeta_3 = e^{2\pi i/3} = (-1 + \sqrt{3}i)/2$, since these are the roots of $f(X)$. We have $E = \mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i)$ and so $[E : \mathbb{Q}] = [E : \mathbb{Q}(\sqrt[3]{2})][\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 2 \cdot 3 = 6$.

4.8 Remark Let E_1/F and E_2/F be two extension fields of F and let $a \in E_1$. Moreover, let $\varphi \in \mathrm{Hom}_{F-\mathsf{alg}}(E_1, E_2)$. If $f(X) := \min_{a/F}(X) = X^n + \alpha_{n-1}X^{n-1} + \cdots + \alpha_0$, then

$$\begin{aligned} 0 = \varphi(0) &= \varphi(a^n + \alpha_{n-1}a^{n-1} + \cdots + \alpha_0) \\ &= \varphi(a)^n + \alpha_{n-1}\varphi(a)^{n-1} + \cdots + \alpha_0 = f(\varphi(a)) \,. \end{aligned}$$

Therefore, $\varphi(a) \in E_2$ is also a root of f. In particular, if there exists an irreducible $f \in F[X]$ which has a root in E_1 but not in E_2, then $\mathrm{Hom}_{F-\mathsf{alg}}(E_1, E_2) = \emptyset$.

4.9 Lemma *Let F and $\tilde{F}$ be fields and let $\varphi \colon F \to \tilde{F}$ be a field isomorphism. Let E/F and $\tilde{E}/\tilde{F}$ be field extensions, let $a \in E$ be algebraic over F with $f(X) := \min_{a/F}(X) = X^n + \alpha_{n-1}X^{n-1} + \cdots + \alpha_0$, and let $\tilde{a} \in \tilde{E}$ be algebraic over $\tilde{F}$ with $\min_{\tilde{a}/\tilde{F}}(X) = \tilde{f}(X) := X^n + \varphi(\alpha_{n-1})X^{n-1} + \cdots + \varphi(\alpha_0)$. Then there exists a unique ring homomorphism $\psi \colon F(a) \to \tilde{F}(\tilde{a})$ such that $\psi(a) = \tilde{a}$ and $\psi(\alpha) = \varphi(\alpha)$ for all $\alpha \in F$. Moreover, ψ is an isomorphisms.*

Proof We extend $\varphi \colon F \to \tilde{F}$ to the ring isomorphism

$$\Phi \colon F[X] \to \tilde{F}[X] \,, \quad \alpha_n X^n + \cdots + \alpha_0 \mapsto \varphi(\alpha_n)X^n + \cdots + \varphi(\alpha_0) \,.$$

Then $\tilde{f} = \Phi(f)$, and Φ induces an isomorphism

$$\bar{\Phi} \colon F[X]/(f) \to \tilde{F}[X]/(\tilde{f}) \,.$$

Composing $\bar{\Phi}$ with the ring isomorphisms $\bar{\rho} \colon F[X]/(f) \cong F(a)$ and $\tilde{\bar{\rho}} \colon \tilde{F}[X]/(\tilde{f}) \cong \tilde{F}(\tilde{a})$ from Theorem 2.4, we obtain a ring isomorphism with the desired properties.

Since each element in $F(a)$ is an F-linear combination of the elements $1, a, \ldots, a^{n-1}$, there exists at most one ring homomorphism $\psi \colon F(a) \to \tilde{F}(\tilde{a})$ such that $\psi(a) = \tilde{a}$ and $\psi(\alpha) = \varphi(\alpha)$ for all $\alpha \in F$. □

4.10 Corollary *Let E/F and $\tilde{E}/F$ be field extensions with $E = F(a)$ for some $a \in E$ which is algebraic over F. Let $f := \min_{a/F} \in F[X]$ and let $\tilde{a}_1, \ldots, \tilde{a}_n$ be the roots of f in $\tilde{E}$ (without repetition). Then, there exist exactly n F-algebra homomorphisms $\varphi_1, \ldots, \varphi_n : E \to \tilde{E}$ and they are given by $\varphi_i(a) = \tilde{a}_i$, for $i = 1, \ldots, n$.*

Proof This is immediate from Lemma 4.9 and Remark 4.8. □

4.11 Theorem *Let $F, \tilde{F}$ be fields, let $\varphi \colon F \to \tilde{F}$ be a field isomorphism, let $f(X) = \alpha_n X^n + \cdots + \alpha_1 X + \alpha_0 \in F[X]$ with $\alpha_n \neq 0$ and $n > 0$, and set $\tilde{f}(X) := \varphi(\alpha_n)X^n + \cdots + \varphi(\alpha_0) \in \tilde{F}[X]$. If E is a splitting field of f over F and if $\tilde{E}$ is a splitting field of $\tilde{f}$ over $\tilde{F}$, then there exists a field isomorphism*

$\psi\colon E \to \tilde{E}$ *such that* $\psi(\alpha) = \varphi(\alpha)$ *for all* $\alpha \in F$. *In particular, if* $F = \tilde{F}$ *and* $\varphi = \mathrm{id}_F$, *we obtain that any two splitting fields of* f *over* F *are isomorphic as* F*-algebras.*

Proof We proof the assertion by induction on $m := [E : F]$. If $m = 1$, then $f(X) = \alpha_n(X - r_1)\cdots(X - r_n)$ with $r_1, \ldots, r_n \in F$, $\tilde{f}(X) = \varphi(\alpha_n)(X - \varphi(r_1))\cdots(X - \varphi(r_n))$, and $\tilde{E} = \tilde{F}$. Hence, we may take $\psi := \varphi$. Now let $m = [E : F] \geqslant 2$. Then $f(X)$ has an irreducible factor $p(X)$ with $\deg(p) \geqslant 2$, and $\tilde{p} := \Phi(p) \in \tilde{F}[X]$ is an irreducible factor of $\tilde{f}$ (with Φ as in the proof of Lemma 4.9). With f also p decomposes into linear factors in $E[X]$ and there exists $a \in E$ with $p(a) = 0$. Similarly there exists $\tilde{a} \in \tilde{E}$ such that $\tilde{p}(\tilde{a}) = 0$. By Lemma 4.9, there exists a field isomorphism $\sigma\colon F(a) \to \tilde{F}(\tilde{a})$ such that $\sigma(\alpha) = \varphi(\alpha)$ for all $\alpha \in F$. Obviously, E (resp. $\tilde{E}$) is still a splitting field of $f(X)$ (resp. $\tilde{f}$) over $F(a)$ (resp. $\tilde{F}(\tilde{a})$). By induction we obtain a field isomorphism $\psi\colon E \to \tilde{E}$ such that $\psi(\beta) = \sigma(\beta)$ for each $\beta \in F(a)$. Hence, $\psi(\alpha) = \varphi(\alpha)$ for each $\alpha \in F$. $\square$

4.12 Definition An algebraic field extension E/F is called *normal*, if each irreducible $f \in F[X]$ which has a root in E splits completely into linear factors in $E[X]$.

4.13 Theorem *Let* E/F *be a finite field extension. The following are equivalent:*

(i) E/F *is normal.*

(ii) E *is a splitting field of some* $f \in F[X]$ *over* F.

(iii) *For each field extension* E'/E *and each* F*-algebra homomorphism* $\varphi\colon E \to E'$, *one has* $\varphi(E) \subseteq E$.

Proof (i) $\Rightarrow$ (ii): Let $a_1, \ldots, a_n \in E$ with $E = F(a_1, \ldots, a_n)$ and set $f(X) := \min_{a_1/F}(X)\cdots\min_{a_n/F}(X)$. Since E/F is normal, f decomposes in $E[X]$ into linear factors. Moreover, E is obtained by adjoining the roots of f in E to F. Hence, E is a splitting field of f over F.

(ii) $\Rightarrow$ (iii): Let $f \in F[X]$ be as in (ii) and write

$$f(X) = \alpha_(X - a_1)\cdots(X - a_n)$$

with $\alpha \in F$ and $a_1, \ldots, a_n \in E$. Then $E = F(a_1, \ldots, a_n)$. Let E'/E be a field extension and let $\varphi \in \mathrm{Hom}_{F\text{-}\mathsf{alg}}(E, E')$. By Remark 4.8, for each $i = 1, \ldots, n$, there exists $\pi(i) \in \{1, \ldots, n\}$ such that $\varphi(a_i) = a_{\pi(i)}$. In particular, $\varphi(E) = \varphi(F(a_1, \ldots, a_n)) \subseteq F(a_1, \ldots, a_n) = E$.

(iii) $\Rightarrow$ (i): Let $a \in E$ and let $f := \min_{a/F}(X) \in F[X]$. Let $a_1, \ldots, a_n \in E$ such that $E = F(a_1, \ldots, a_n)$ and let E' be a splitting field of

$$g(X) := f(X) \cdot \min_{a_1/F}(X)\cdots\min_{a_n/F}(X)$$

over E. Let $b \in E'$ with $f(b) = 0$. Then, by Lemma 4.9, there exists an F-algebra isomorphism $\sigma\colon F(a) \to F(b)$ with $\sigma(a) = b$. Note that E' is also a splitting field of $g(X)$ over F, $F(a)$, and $F(b)$. By Theorem 4.11, σ can be extended to an F-algebra isomorphism $\psi\colon E' \to E'$. Let $\varphi\colon E \to E'$ be the restriction of ψ to E. By (iii) we then have $b = \psi(a) \in \psi(E) \subseteq E$, so that $f(X)$ splits in $E[X]$ into linear factors. □

4.14 Corollary *Let E/F be a finite field extension. Then there exists a finite field extension E'/E such that E'/F is normal.*

Proof Let $a_1, \ldots, a_n \in E$ such that $E = F(a_1, \ldots, a_n)$ and let E' be a splitting field of $f(X) := \min_{a_1/F}(X) \cdots \min_{a_n/F}(X)$ over E. Then E' is also a splitting field of f over F. Therefore, E'/F is normal by Theorem 4.13. □

4.15 Example $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not normal, since $\sqrt[3]{2} \in \mathbb{Q}(\sqrt[3]{2})$ is a root of the irreducible polynomial $X^3 - 2 \in \mathbb{Q}[X]$, but the other roots $\zeta_3 \sqrt[3]{2}$ and $\zeta_3 \sqrt[3]{2}$ are not in $\mathbb{Q}(\sqrt[3]{2})$. However, $\mathbb{Q}(\sqrt[3]{2}, \sqrt{3}i)$ is normal over $\mathbb{Q}$, since it is a splitting field of $X^3 - 2$ over $\mathbb{Q}$ (see Example 4.7).

5 Separable Field Extensions

5.1 Definition Let F be a field.

(a) An irreducible polynomial $f \in F[X]$ is called *separable*, if f has $\deg(f)$ distinct roots in a splitting field of f over F. An arbitrary polynomial $f \in F[X] \smallsetminus F$ is called *separable over* F, if every irreducible factor of f in $F[X]$ is separable.

(b) Let E/F be an algebraic field extension. An element $a \in E$ is called *separable over* F, if $\min_{a/F}(X)$ is separable. The field extension E/F is called *separable*, if every element $a \in E$ is separable over F. We call F *perfect*, if every algebraic extension E/F is separable.

(c) We call the F-linear map

$$\frac{d}{dX}\colon F[X] \to F[X]\,, \quad X^n \mapsto nX^{n-1}\,,$$

the *formal derivation* and $f'(X) := \frac{d}{dX}(f(X))$ the *formal derivative* of $f(X) \in F[X]$. It is straight forward to verify that $(f+g)' = f' + g'$, $(\alpha f)' = \alpha(f')$, $(f \cdot g)' = f' \cdot g + f \cdot g'$, and $(f \circ g)' = (f' \circ g) \cdot g'$, for $\alpha \in F$ and $f, g \in F[X]$. Note that $\frac{d}{dX}(X^p) = 0$ if F has characteristic p.

5.2 Lemma *Let F be a field and let $f \in F[X]$ be irreducible. Then the following are equivalent:*

(i) *f is separable.*

(ii) *For every field extension E/F, $f(X)$ decomposes in $E[X]$ into distinct irreducible factors.*

(iii) *f and f' are coprime in $F[X]$.*

(iv *$f' \neq 0$.*

Proof (i) $\Rightarrow$ (ii): Let E/F be a field extension and suppose that $f(X) = p(X)^2 g(X)$ with $p, g \in E[X]$ and p irreducible in $E[X]$. Then $f(X)$ has less than $\deg(f)$ distinct roots in a splitting field of f over E, and then also in a splitting field over F, which is a contradiction.

(ii) $\Rightarrow$ (iii): Let E be a splitting field of f over F and let $f(X) = \alpha(X - r_1)\cdots(X - r_n)$ be a decomposition in $E[X]$. We have (by induction on n)

$$f'(X) = \alpha \sum_{i=1}^{n} (X - r_1)\cdots(X - r_{i-1})(X - r_{i+1})\cdots(X - r_n)\,.$$

Assume that f and f' have a common irreducible factor $p \in F[X]$. Then $p(r_i) = 0$ for some $i = 1\ldots, n$ and we obtain

$$0 = f'(r_i) = (r_i - r_1)\cdots(r_i - r_{i-1})(r_i - r_{i+1})\cdots(r_i - r_n)\,.$$

This implies $r_i = r_j$ for some $j \in \{1, \ldots, n\}$ with $i \neq j$, which contradicts (ii).

(iii) $\Rightarrow$ (iv): Assume that $f' = 0$. Then f is a common divisor of f and f'.

(iv) $\Rightarrow$ (i): Let E be a splitting field of f over F and let $f = \alpha(X - r_1)\cdots(X - r_n)$ with $\alpha \in F$ and $r_1, \ldots, r_n \in E$. Assume that $r_i = r_j$ for some $i \neq j$ from $\{1, \ldots, n\}$. Then $(X - r_i)$ is a common divisor of f and f' in $E[X]$, hence $f'(r_i) = 0$. But this implies that $f = \min_{r_i/F}(X) \mid f'$. This is a contradiction, since $\deg(f') < \deg(f)$ and $f' \neq 0$. $\square$

5.3 Proposition *Let F be a field and let $f \in F[X]$.*

(a) If $\mathrm{char}(F) = 0$, then one has: $f' = 0 \iff f \in F$.

(b) If $\mathrm{char}(F) = p > 0$, then one has: $f' = 0 \iff f(X) = g(X^p)$ for some $g \in F[X]$.

Proof Let $f = \sum_{i=0}^{n} \alpha_i X^i$, so $f' = \sum_{i=1}^{n} \alpha_i i X^{i-1}$. Therefore, $f' = 0$ if and only if $\alpha_i i = 0$ for all $i \geqslant 1$. If $\mathrm{char}(F) = 0$, then the last statement is equivalent to $\alpha_i = 0$ for all $i \geqslant 1$. In the case of positive characteristic it is equivalent to $\alpha_i = 0$ for all $i \geqslant 1$ which are not divisible by p, and this implies

$$f = \alpha_0 + \alpha_p X^p + \alpha_{2p} X^{2p} + \cdots = g(X^p)$$

with $g = \alpha_0 + \alpha_p X + \alpha_{2p} X^2 + \cdots$. Conversely, if $g \in F[X]$ is arbitrary, then $g(X^p) \in F[X]$ has vanishing derivative by Definition 5.1. $\square$

5.4 Corollary *Every field F of characteristic 0 is perfect.*

Proof By Lemma 5.2 and Proposition 5.3 every $f \in F[X]$ which is not constant is separable over F. $\square$

5.5 Definition Let F be a field of characteristic $p > 0$. Note that for $\alpha, \beta \in F$ we have

$$(\alpha+\beta)^p = \sum_{i=0}^{p} \binom{p}{i} \alpha^i + \beta^{p-i} = \alpha^p \beta^p \,,$$

since $\binom{p}{i}$ is divisible by p for all $i = 1, \ldots, p-1$. Therefore, the map $\varphi\colon F \to F$, $\alpha \mapsto \alpha^p$, is a ring homomorphism which we call the *Frobenius homomorphism* of F.

5.6 Theorem *Let F be a field of characteristic $p > 0$. Then F is perfect if and only if the Frobenius homomorphism is surjective.*

Proof $\Rightarrow$: Let $\alpha \in F$. We have to show that $\alpha = \beta^p$ for some $\beta \in F$. Let E be a splitting field of $f(X) := X^p - \alpha$ over F, and let $b \in E$ be a zero of $f(X)$, i.e., $\alpha = b^p$. Then $\min_{b/F}(X)$ divides $X^p - \alpha = X^p - b^p = (X-b)^p$. Therefore, $\min_{b/F}(X) = (X-b)^r$ for some $r \in \{1, \ldots, p\}$. But since b is separable over F, $r = 1$ and $b \in F$.

$\Leftarrow$: Let E/F be an algebraic field extension. Assume that some $a \in E$ is not separable over F. Then $\min'_{a/F}(X) = 0$ and so $\min_{a/F}(X) = \alpha_0 + \alpha_1 X^p + \cdots + X^{np}$. Let $\beta_i \in F$ with $\alpha_i = \beta_i^p$ for each $i = 0, \ldots, n-1$. Then

$$\begin{aligned}\min_{a/F}(X) = \beta_0^p + \beta_1^p X^p + \cdots \beta_{n-1}^p X^{(n-1)p} + X^{np} = \\ (\beta_0 + \beta_1 X + \cdots + \beta_{n-1} X^{n-1} + X^n)^p\end{aligned}$$

contradicting the irreducibility of $\min_{a/F}(X)$. $\square$

5.7 Corollary *Every finite field F is perfect.*

Proof The Frobenius homomorphism $\varphi\colon F \to F$ is injective and then also surjective, since F is finite. $\square$

5.8 Example Let $F := \mathbb{F}_2(T)$ be the function field over the field with 2 elements in the indeterminate T. Then $f(X) := X^2 - T \in F[X]$ is irreducible. In fact, if not, then there exists $\alpha \in F$ with $T = \alpha^2 = p^2/q^2$ for some $p, q \in \mathbb{F}_2[X]$, $q \neq 0$. But then, $Tq(T)^2 = p(T)^2$, which is impossible, since $Tq(T)^2$ has odd degree and $p(T)^2$ has even degree. Therefore, the splitting field of $f(X)$ over F is not separable, i.e., $F(\sqrt{T})/F$ is not separable.

5.9 Theorem *Let F be a field and let $G \leqslant F^\times$ be a finite subgroup of the unit group $F^\times := F \smallsetminus \{0\}$ of F. Then G is cyclic. In particular, if F is finite, then $F^\times$ is cyclic.*

Proof Let $|G| = p_1^{\alpha_1} \cdots p_t^{\alpha_t}$ be the prime factor decomposition of $|G|$ and let $P_i \leqslant G$ be the Sylow p_i-subgroups of G. Since G is Abelian, we have $G \cong P_1 \times \cdots \times P_t$, and by the Chinese Remainder Theorem it suffices to show that each P_i is cyclic. Assume that P_i is not cyclic. Then $g^{p_i^{\alpha_i - 1}} = 1$ for each $g \in P_i$. But then the polynomial $X^{p_i^{\alpha_i - 1}} - 1 \in F[X]$ has at least $p_i^{\alpha_i}$ different roots in F, which is a contradiction. □

5.10 Theorem *Let E/F be a finite separable field extension. Then there exists* a primitive element *for E/F, i.e. an element $a \in E$ such that $E = F(a)$.*

Proof If F is finite, also E is finite and $E = F(a)$ for every $a \in E^\times$ with $\langle a \rangle = E^\times$. Therefore, we assume that $|F| = \infty$ and write $E = F(x_1, \ldots, x_n)$ with $x_1, \ldots, x_n \in E$ and minimal $n \in \mathbb{N}$. Assume that $n > 1$. Set $a := x_1$ and $b := x_2$, let E_1 be a splitting field of $\min_{a/F} \cdot \min_{b/F}$ over E, and write

$$m_{a/F}(X) = \prod_{i=1}^{r} (X - a_i), \quad m_{b/F}(X) = \prod_{j=1}^{s} (X - b_j),$$

with $a = a_1, \ldots, a_r, b = b_1, \ldots, b_s \in E_1$. Since b is separable over F, the elements $b_1, \ldots, b_s$ are pairwise distinct, and there exists $\gamma \in F$ such that $\gamma \neq \frac{a_i - a}{b - b_j}$ for each $i = 1, \ldots, r$ and $j = 2, \ldots, s$. Set $c := a + \gamma b = a_1 + \gamma b_1$. Then $F(c) \subseteq F(a, b)$. Since $\min_{a/F}(c - \gamma b) = 0$, the polynomial $(X - b)$ divides $\min_{a/F}(c - \gamma X)$ and $\min_{b/F}(X)$ in $F(a, b)[X]$. By the choice of γ, $c - \gamma b_j = a + \gamma(b - b_j) \neq a_i$ for any $i = 1, \ldots, r$ and $j = 1, \ldots, s$. Therefore, $\min_{a/F}(c - \gamma b_j) \neq 0$ for all $j = 2, \ldots, s$, and $X - b$ is a gcd of $\min_{b/F}(X)$ and $\min_{a/F}(c - \gamma X)$ in $F(a, b)[X]$. But $\min_{b/F}(X)$ and $\min_{a/F}(c - \gamma X)$ are contained in $F(c)[X]$, and so $X - b \in F(c)[X]$ (use the Euclidean algorithm or see Exercise 39). Therefore, $b \in F(c)$ and also $a = c - \gamma b \in F(c)$. This implies $F(a, b) \subseteq F(c)$ and $E = F(c, x_3, \ldots, x_n)$, contradicting the minimality of n. □

5.11 Lemma *Let E/F be an algebraic field extension, assume that* $\mathrm{char}(F) =$ *$p > 0$, and let $a \in E$. Then $F(a^p) \subseteq F(a)$, and one have:*

(a) *If a is separable over F, then $a \in F(a^p)$ and $F(a) = F(a^p)$.*

(b) *If a is not separable over F, then $[F(a) : F(a^p)] = p$.*

Proof Obviously $F(a^p) \subseteq F(a)$.

(a) Assume that a is separable over F. Set $E' := F(a^p) \subseteq F(a)$. Since $\min_{a/E'} \mid \min_{a/F}$ and since $\min_{a/F}$ is separable, also $\min_{a/E'}$ is separable. Clearly, $\min_{a/E'}(X) \mid X^p - a^p \in F(a)[X]$. This implies that $\min_{a/E'} = X - a$ and $a \in E'$.

(b) Assume that a is not separable over F. Then $\min'_{a/F}(X) = 0$ by Lemma 5.2, and $\min_{a/F}(X) = g(X^p)$ for some monic irreducible $g \in F[X]$

by Proposition 5.3. Since $g(a^p) = \min_{a/F}(a) = 0$, we have $g = \min_{a^p/F}(X)$. This implies that

$$\begin{aligned}[F(a) : F(a^p)][F(a^p) : F] = [F(a) : F] &= \deg(\min_{a/F}) \\ = p\deg(\min_{a^p/F}) &= p[F(a^p) : F]\,,\end{aligned}$$

and so $[F(a) : F(a^p)] = p$. □

5.12 Theorem *Let E/F be an algebraic field extension and let $a_1, \ldots, a_n \in E$. Assume that a_i is separable over $F(a_1, \ldots, a_{i-1})$ for each $i = 1, \ldots, n$. Then $F(a_1, \ldots, a_n)/F$ is separable.*

Proof We may assume that $\mathrm{char}(F) = p > 0$. Set $F_0 := F$ and $F_i := F_{i-1}(a_i)$ for $i = 1, \ldots, n$. Then $F_i = F_{i-1}(a_i^p)$ for each $i = 1, \ldots, n$ by Lemma 5.10. Therefore, each element of

$$F_n = F_{n-1}(a_n^p) = F_{n-2}(a_{n-1}^p, a_n^p) = \cdots = F(a_1^p, \ldots, a_n^p)$$

is an F-linear combination of p-powers of elements in F_n. Let $b \in F_n$ and set $k := \deg(\min_{b/F}(X)) = [F(b) : F]$. We will show that $F(b) = F(b^p)$ which implies that b is separable over F by Lemma 5.11. For that purpose we extend the F-basis $b_1 := 1, b_2 := b, \ldots, b_k := b^{k-1}$ of $F(b)$ to an F-basis $b_1, \ldots, b_l$ of F_n. Let $x \in F_n$ be arbitrary. Then there exist $\xi_1, \ldots, \xi_l \in F$ with $x = \xi_1 b_1 + \cdots + \xi_l b_l$ and we have $x^p = \xi_1^p b_1^p + \cdots + \xi_l^p b_l^p$. Since every element of F_n is an F-linear combination of elements of the form x^p, the elements $b_1^p, \ldots, b_l^p$ span F_n as an F-vector space, Therefore they form an F-basis of F_n. In particular, $1, b^p, \ldots, b^{p(k-1)}$ are F-linearly independent and we obtain $k = [F(b) : F] \geqslant [F(b^p) : F] \geqslant k$, which implies $[F(b) : F] = [F(b^p) : F]$ and $F(b) = F(b^p)$. □

5.13 Corollary *Let E/F be an algebraic field extension. If $E = F(a_1, \ldots, a_n)$ for elements $a_1, \ldots, a_n \in E$ which are separable over F, then E/F is separable.*

Proof If a_i is separable over F, so also over $F(a_1, \ldots, a_{i-1})$, since $\min_{a_i/F(a_1,\ldots,a_{i-1})} \mid \min_{a_i/F}$. Now Theorem 5.12 applies. □

6 Galois Extensions

6.1 Definition For a field extension E/F we denote by $\mathrm{Aut}(E/F)$ the set of F-algebra isomorphisms $\varphi\colon E \to E$. Note that $\mathrm{Aut}(E/F)$ is a group under composition.

6.2 Remark In the above situation, $\mathrm{Aut}(E/F)$ is contained in $\mathrm{Hom}_F(E,E)$, the group of F-linear homomorphisms from E to E. Note that $\mathrm{Hom}_F(E,E)$ is an E-vector space by setting $(x \cdot f)(y) := x(f(y))$ for $f \in \mathrm{Hom}_F(E,E)$ and $x, y \in E$. It is not difficult to see that if E/F has finite degree n, then $\dim_E(\mathrm{Hom}_F(E,E)) = n$ (see Exercise 42).

6.3 Lemma (Dedekind) *For every field extension E/F, the set $\mathrm{Aut}(E/F)$ is E-linearly independent in the E-vector space $\mathrm{Hom}_F(E,E)$. In particular, if $[E:F]$ is finite, then $|\mathrm{Aut}(E/F)| \leqslant [E:F]$.*

Proof Assume that $\mathrm{Aut}(E/F)$ is not E-linearly independent and let $a_1\varphi_1 + \cdots + a_n\varphi_n = 0$ be a non-trivial relation with $a_1, \ldots a_n \in E$ and $\varphi_1, \ldots, \varphi_n \in \mathrm{Aut}(E/F)$ such that n is minimal. Then $n \geqslant 2$, $a_1 \neq 0$, and there exists $y \in E$ with $\varphi_1(y) \neq \varphi_n(y)$. We have for each $x \in E$:

$$0 = a_1\varphi_1(xy) + \cdots + a_n\varphi_n(xy) = a_1\varphi_1(y)\varphi_1(x) + \cdots + a_n\varphi_n(y)\varphi_n(x)$$

and

$$0 = \varphi_n(y)(a_1\varphi_1(x) + \cdots + a_n\varphi_n(x)) = a_1\varphi_n(y)\varphi_1(x) + \cdots + a_n\varphi_n(y)\varphi_n(x)\,.$$

Subtraction yields a shorter relation

$$0 = (b_1\varphi_1 + \cdots + b_{n-1}\varphi_{n-1})(x)$$

for all $x \in E$ with $b_i := a_i(\varphi(y) - \varphi_n(y))$ for $i = 1, \ldots, n-1$ and $b_1 \neq 0$. This contradicts the minimality of n. $\square$

6.4 Remark Let E/F be a field extension and let $X \subseteq \mathrm{Aut}(E/F)$ be a subset. Then the set

$$\mathrm{Fix}_E X := \{a \in E \mid \varphi(a) = a \text{ for all } \varphi \in X\}$$

is a subfield of E containing F. Conversely, if F' is an intermediate field in E/F, then $\mathrm{Aut}(E/F')$ is a subgroup of $\mathrm{Aut}(E/F)$. Therefore we obtain maps between the set of intermediate fields F' in E/F and the set of subgroups H of $G := \mathrm{Aut}(E/F)$:

$$\begin{aligned} \mathrm{Fix}_E \colon H &\mapsto \mathrm{Fix}_E(H)\,, \\ \mathrm{Aut}(E/-) \colon F' &\mapsto \mathrm{Aut}(E/F')\,. \end{aligned}$$

Obviously, we have

$$F' \subseteq \mathrm{Fix}_E(\mathrm{Aut}(E/F'))\,, \quad H \leqslant \mathrm{Aut}(E/Fix_E(H))\,, \tag{6.4.a}$$

for each intermediate field F' of E/F and each subgroup $H \leqslant G$. Moreover, if $F' \subseteq F''$ are intermediate fields of E/F and if $H \leqslant K$ are subgroups of G, then

$$\mathrm{Aut}(E/F'') \leqslant \mathrm{Aut}(E/F') \quad \text{and} \quad \mathrm{Fix}_E(K) \subseteq \mathrm{Fix}_E(H)\,. \tag{6.4.b}$$

Note that (6.4.a) and (6.4.b) imply

$$\mathrm{Fix}_E(\mathrm{Aut}(E/Fix_E(H))) = \mathrm{Fix}_E(H) \tag{6.4.c}$$

and

$$\mathrm{Aut}(E/Fix_E(\mathrm{Aut}(E/F))) = \mathrm{Aut}(E/F)\,. \tag{6.4.d}$$

6.5 Definition A finite field extension E/F is called a *Galois extension* if $F = \mathrm{Fix}_E(\mathrm{Aut}(E/F))$. In this case, we call $\mathrm{Aut}(E/F)$ the *Galois group* of E/F and denote it by $\mathrm{Gal}(E/F)$.

6.6 Example (a) $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is a Galois extension, since $\mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = \{1,\tau\}$ with $\tau(\sqrt{2}) = -\sqrt{2}$, and since $\mathrm{Fix}_{\mathbb{Q}(\sqrt{2})}(\{1,\tau\}) = \{a+b\sqrt{2} \mid a,b \in \mathbb{Q},\ a+b\sqrt{2} = a-b\sqrt{2}\} = \mathbb{Q}$.

(b) $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not a Galois extension, since $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = \{1\}$ and since $\mathrm{Fix}_{\mathbb{Q}(\sqrt[3]{2})}(\{1\}) = \mathbb{Q}(\sqrt[3]{2}) \neq \mathbb{Q}$.

(c) $\mathbb{C}/\mathbb{R}$ is a Galois extension, since $\mathrm{Aut}(\mathbb{C}/\mathbb{R}) = \{1,\bar{\cdot}\}$ and $\mathrm{Fix}_{\mathbb{C}}(\{1,\bar{\cdot}\}) = \mathbb{R}$.

6.7 Theorem *For a finite field extension E/F the following are equivalent:*

(i) *E/F is a Galois extension.*

(ii) *E/F is separable and normal.*

(iii) *E is a splitting field of some polynomial $f \in F[X]$ which is separable over F.*

(iv) *There exists $a \in E$ such that $E = F(a)$ and $\min_{a/F}(X)$ has $[E:F]$ distinct roots in E.*

(v) *$|\mathrm{Aut}(E/F)| = [E:F]$.*

Proof (i)$\Rightarrow$(ii): First we show that E/F is separable. Let $a \in E$ and let $\{\sigma(a) \mid \sigma \in \mathrm{Aut}(E/F)\} = \{a_1,\ldots,a_n\}$ such that $a = a_1,\ldots,a_n$ are pairwise distinct. Set $f(X) := (X-a_1)\cdots(X-a_n) \in E[X]$ and write $f(X) = X^n + b_{n-1}X^{n-1} + \cdots + b_0$ with $b_0,\ldots,b_{n-1} \in E$. For $\sigma \in \mathrm{Aut}(E/F)$ denote by $\tilde{\sigma}\colon E[X] \to E[X]$ the induced F-algebra isomorphism. Since each $\sigma \in \mathrm{Aut}(E/F)$ just permutes $a_1,\ldots,a_n$, we have $\tilde{\sigma}(f) = f$ and therefore, $\sigma(b_i) = b_i$ for all $i = 1,\ldots,n-1$. By (i), this implies $b_0,\ldots,b_{n-1} \in F$ and $\min_{a/F}(X) \mid f$. Therefore, a is separable over F.

By Theorem 5.10, there exists $a \in E$ such that $E = F(a)$. With the above considerations for this a, we see that E is a splitting field of $\min_{a/F}$ over F, and therefore, E/F is normal by Theorem 4.13.

(ii)$\Rightarrow$(iii): Since E/F is normal, it is a splitting field for some monic polynomial $f(X) \in F[X]$. Since E/F is separable, each of the roots of $f(X)$ in E is separable over F. Therefore, their minimal polynomials (which are just the irreducible factors of $f(X)$ in $F[X]$) are separable. By definition, also $f(X)$ is separable over F.

(iii)$\Rightarrow$(iv): E/F is normal by Theorem 4.13. All the roots $a_1,\ldots,a_n \in E$ of f are separable over F and $E = F(a_1,\ldots,a_n)$. Therefore, E/F is separable by Corollary 5.12. By Theorem 5.10, there exists $a \in E$ with $E = F(a)$. Since E/F is normal, $\min_{a/F}(X)$ splits completely into linear factors in $E[X]$. They are pairwise distinct, since a is separable over F. Therefore, $\min_{a/F}(X)$ has $\deg(\min_{a/F}) = [E : F]$ pairwise distinct roots in E.

(iv)⇒(v): Let $a = a_1, \ldots, a_n \in E$ be the distinct roots of $\min_{a/F}(X)$. Note that $n = [E : F]$. By Corollary 4.10, there exist exactly n F-algebra homomorphisms $\varphi_1, \ldots, \varphi_n \colon E \to E$, given by $\varphi_i(a) = a_i$, for $i = 1, \ldots, n$. Since they are injective, they are also surjective, and $|\mathrm{Aut}(E/F)| = n = [E : F]$.

(v)⇒(i): (6.4.d), Lemma 6.3, and (6.4.a) imply

$$\begin{aligned}|\mathrm{Aut}(E/F)| = |\mathrm{Aut}(E/\mathrm{Fix}_E(\mathrm{Aut}(E/F)))| \leqslant \\ \leqslant [E : \mathrm{Fix}_E(\mathrm{Aut}(E/F))] \leqslant [E : F] = |\mathrm{Aut}(E/F)|\,.\end{aligned}$$

Therefore, $[E : \mathrm{Fix}_E(\mathrm{Aut}(E/F))] = [E : F]$ and $F = \mathrm{Fix}_E(\mathrm{Aut}(E/F))$. □

6.8 Corollary *Let E/F be a Galois extension and let $F \subseteq F' \subseteq E$ be an intermediate field. Then also E/F' is a Galois extension.*

Proof This is immediate from part (iv) of Theorem 6.6. □

6.9 Lemma *Let E/F be a field extension and let $H \leqslant \mathrm{Aut}(E/F)$ be a finite subgroup. Then $[E : \mathrm{Fix}_E(H)] \leqslant |H|$.*

Proof Assume that $n := |H| < [E : \mathrm{Fix}_E(H)]$, choose elements $a_1, \ldots, a_{n+1} \in E$ which are linearly independent over $\mathrm{Fix}_E(H)$, and write $H = \{\varphi_1, \ldots, \varphi_n\}$. Then there exists a solution $(x_1, \ldots, x_{n+1}) \in E^{n+1} \smallsetminus \{0\}$ of the linear equation system

$$\sum_{j=1}^{n+1} \varphi_i^{-1}(a_j) x_j = 0\,, \quad i = 1, \ldots, n.$$

Let $l \in \{1, \ldots, n+1\}$ with $x_l \neq 0$. By Lemma 6.3, there exists $y \in E$ with $\varphi_1(y) + \cdots + \varphi_n(y) \neq 0$. Then also $(z_1, \ldots, z_{n+1}) := y x_l^{-1}(x_1, \ldots, x_{n+1})$ is a solution of the above equation system with $z_l = y \neq 0$. Applying φ_i to the i-th equation we obtain

$$\sum_{j=1}^{n+1} \varphi_i(z_j) a_j = 0\,, \quad \text{for } i = 1, \ldots, n.$$

Addition of these equations yields

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{n+1} \varphi_i(z_j) a_j = \sum_{j=1}^{n+1} \Big(\sum_{i=1}^{n} \varphi_i(z_j) \Big) a_j\,.$$

But $\sum_{i=1}^n \varphi_i(z_j) \in \mathrm{Fix}_E(H)$, for all $j = 1, \dots, n+1$, since

$$\sigma\Big(\sum_{i=1}^n \varphi_i(z_j)\Big) = \sum_{i=1}^n (\sigma\varphi_i)(z_j)$$

and $\{\sigma\varphi_1, \dots, \sigma\varphi_n\} = \{\varphi_1, \dots, \varphi_n\}$ for all $\sigma \in H$. Since $a_1, \dots, a_{n+1}$ are linearly independent over $\mathrm{Fix}_E(H)$, we obtain $\sum_{i=1}^n \varphi_i(z_j) = 0$ for all $j = 1, \dots, n+1$, in particular for $j = l$ and $z_l = y$. This is a contradiction. □

6.10 Corollary *Let E/F be a field extension and let $H \leqslant \mathrm{Aut}(E/F)$ be a finite subgroup. Then $H = \mathrm{Aut}(E/\mathrm{Fix}_E(H))$ and $E/\mathrm{Fix}_E(H)$ is a Galois extension with $\mathrm{Gal}(E/\mathrm{Fix}_E(H)) = H$.*

Proof By (6.4.a), Lemma 6.3, and Lemma 6.9, be have

$$|H| \leqslant |\mathrm{Aut}(E/\mathrm{Fix}_E(H))| \leqslant [E : \mathrm{Fix}_E(H)] \leqslant |H|\,,$$

and we have equality everywhere. In particular, $H = \mathrm{Aut}(E/\mathrm{Fix}_E(H))$, and $[E : \mathrm{Fix}_E(H)] = |H| < \infty$, so that $E/\mathrm{Fix}_E(H)$ is a Galois extension. □

6.11 Theorem (Main theorem of Galois theory) *Let E/F be a Galois extension. Set $G := \mathrm{Gal}(E/F)$. Then the maps Fix_E and $\mathrm{Aut}(E/-)$ are inverse bijections between the set of subgroups H of G and the set of intermediate fields F' of E/F. If F' and H are in correspondence, then $[E : F'] = |H|$ and $[F' : F] = [G : H]$.*

Proof Let F' be an intermediate field of E/F. Then E/F' is a Galois extension by Corollary 6.8, and so $F' = \mathrm{Fix}_E(\mathrm{Aut}(E/F'))$. Moreover, for each $H \leqslant G$ we have $H = \mathrm{Aut}(E/\mathrm{Fix}_E(H))$ by Corollary 6.10. If H and F' correspond to each other, then $|H| = |\mathrm{Aut}(E/F')| = [E : F']$ by Corollary 6.7. Moreover,

$$[F' : F] = \frac{[E : F]}{[E : F']} = \frac{|G|}{|H|} = [G : H]\,.$$

□

6.12 Remark Let E/F be a Galois extension and set $G := \mathrm{Gal}(E/F)$. Let $H \leqslant G$ and $\sigma \in G$. Then

$$\mathrm{Fix}_E(\sigma H \sigma^{-1}) = \sigma(\mathrm{Fix}_E(H))\,.$$

In fact, for $a \in E$ we have:

$$\begin{aligned} a \in \mathrm{Fix}_E(\sigma H \sigma^{-1}) &\iff \sigma\tau\sigma^{-1}(a) = a \text{ for all } \tau \in H \\ &\iff \tau(\sigma^{-1}(a)) = \sigma^{-1}(a) \text{ for all } \tau \in H \\ &\iff \sigma^{-1}(a) \in \mathrm{Fix}_E(H) \\ &\iff a \in \sigma(\mathrm{Fix}_E(H))\,. \end{aligned}$$

6.13 Theorem *Let E/F be a Galois extension, let $H \leqslant G := \mathrm{Aut}(E/F)$, and set $F' := \mathrm{Fix}_E(H)$. Then the following are equivalent:*

(i) *F'/F is a Galois extension.*

(ii) *F'/F is normal.*

(iii) *$\sigma(F') \subseteq F'$ for all $\sigma \in G$.*

(iv) *H is a normal subgroup of G.*

Moreover, if (i)–(iv) hold, then the restriction map $r\colon \mathrm{Gal}(E/F) \to \mathrm{Gal}(F'/F)$, $\sigma \mapsto \sigma|_{F'}$, is an epimorphism with kernel H. In particular, r induces an isomorphism

$$\bar{r}\colon G/H \to \mathrm{Gal}(F'/F)\,.$$

Proof (i)$\Rightarrow$(ii): This is immediate from Theorem 6.7.

(ii)$\Rightarrow$(iii): This follows from Theorem 4.13(iii).

(iii)$\Rightarrow$(iv): This follows from Remark 6.12, since

$$\sigma H \sigma^{-1} = \mathrm{Aut}(E/\sigma(F')) \geqslant \mathrm{Aut}(E/F') = H$$

for all $\sigma \in G$ implies $\sigma^{-1} H \sigma \leqslant H$ for all $\sigma \in G$.

(iv)$\Rightarrow$(iii): If $\sigma H \sigma^{-1} = H$ for all $\sigma \in G$, then $\sigma(F') = F'$ for all $\sigma \in G$ by Remark 6.12.

(iii)$\Rightarrow$(i): If (iii) holds, then $r\colon \mathrm{Aut}(E/F) \to \mathrm{Aut}(F'/F)$, $\sigma \mapsto \sigma|_{F'}$, is a well-defined group homomorphism. For $\sigma \in \mathrm{Aut}(E/F)$ we have: $\sigma \in \ker(r) \iff \sigma|_{F'} = \mathrm{id}_{F'} \iff \sigma \in \mathrm{Aut}(E/F')$. Therefore, r induces a monomorphism $\bar{r}\colon \mathrm{Aut}(E/F)/\mathrm{Aut}(E/F') \to \mathrm{Aut}(F'/F)$. But, by Theorem 6.11 and Lemma 6.3, we also have

$$|\mathrm{Aut}(E/F)/\mathrm{Aut}(E/F')| = [G:H] = [F:F'] \geqslant |\mathrm{Aut}(F'/F)|\,.$$

This implies, that we have equality and that $\bar{r}$ and r are surjective. Moreover, since $[F:F'] = |\mathrm{Aut}(F'/F)|$, F'/F is a Galois extension by Theorem 6.7. □

6.14 Definition Let F be a field and let $f(X) \in F[X]$ be a separable irreducible polynomial. We define the *Galois group of f over F* as the Galois group G of a splitting field E of $f(X)$ over F. By Theorem 4.11, two different such splitting fields have isomorphic Galois groups. Note that, by Remark 4.8, G acts on the set $\{r_1, \ldots, r_n\}$ of roots of $f(X)$ in E. Moreover, if $\sigma \in G$ acts as the identity on $\{r_1, \ldots, r_n\}$, then $\sigma = \mathrm{id}_E$, since $E = F(r_1, \ldots, r_n)$. Therefore, we can regard G as a subgroup of the symmetric group $\mathrm{Sym}(r_1, \ldots, r_n) \cong S_n$.

6.15 Theorem *Assume the notation of Definition 6.14. Then G acts transitively on $R = \{r_1, \ldots, r_n\}$.*

Proof Let R_1 be the G-orbit of r_1 and set $R_2 := R \smallsetminus R_1$. We will show that R_2 is empty. Write $f(X) = \alpha(X - r_1)\cdots(X - r_n)$ with $\alpha \in F$. If we define

$$g(X) := \alpha \prod_{r \in R_1} (X - r) \quad \text{and} \quad h(X) := \prod_{r \in R_2} (X - r)\,.$$

then $f = gh$. For $\sigma \in G$, let $\tilde{\sigma}\colon E[X] \to E[X]$ denote the induced ring isomorphism. Then

$$\tilde{\sigma}(g(X)) = \alpha \prod_{r \in R_1} (X - \sigma(r)) = g(X)\,,$$

since $\{\sigma(r) \mid r \in R_1\} = R_1$. Similarly, $\tilde{\sigma}(h(X)) = h(X)$. This implies that $g(X)$ and $h(X)$ have coefficients in $\mathrm{Fix}_E(G) = F$. Since $f(X)$ is irreducible and $g(X)$ is not constant, $h(X) = 1$ and $R_2 = \emptyset$. $\square$

6.16 Example Let $E := \mathbb{Q}(\sqrt[3]{2}, \zeta)$ with $\zeta := e^{2\pi i/3}$. Then E is a splitting field of the separable polynomial $f(X) = X^3 - 2$. Therefore, E/F is a Galois extension with $G := \mathrm{Gal}(E/\mathbb{Q})$ of order $[E : \mathbb{Q}] = 6$. Since $\min_{\sqrt[3]{2}/\mathbb{Q}}(X) = X^3 - 2 = (X - \sqrt[3]{2})(X - \zeta\sqrt[3]{2})(X - \zeta^2\sqrt[3]{2})$ and $\min_{\zeta/\mathbb{Q}}(X) = X^2 + X + 1 = (X - \zeta)(X - \zeta^2)$ (we have $X^3 - 1 = (X - 1)(X^2 + X + 1)$), Remark 4.8 only allows the following 6 possibilities of $\mathbb{Q}$-algebra homomorphisms from E to E:

$$\begin{aligned}
\mathrm{id}\colon &\quad \sqrt[3]{2} \mapsto \sqrt[3]{2}\,, && \zeta \mapsto \zeta\,,\\
\sigma\colon &\quad \sqrt[3]{2} \mapsto \zeta\sqrt[3]{2}\,, && \zeta \mapsto \zeta\,,\\
\sigma^2\colon &\quad \sqrt[3]{2} \mapsto \zeta^2\sqrt[3]{2}\,, && \zeta \mapsto \zeta\,,\\
\tau\colon &\quad \sqrt[3]{2} \mapsto \sqrt[3]{2}\,, && \zeta \mapsto \zeta^2\,,\\
\sigma\tau\colon &\quad \sqrt[3]{2} \mapsto \zeta\sqrt[3]{2}\,, && \zeta \mapsto \zeta^2\,,\\
\sigma^2\tau\colon &\quad \sqrt[3]{2} \mapsto \zeta^2\sqrt[3]{2}\,, && \zeta \mapsto \zeta^2\,.
\end{aligned}$$

Since $|G| = 6$, these must be the elements of G. Note that in this example, G considered as a permutation group of the roots $\{\sqrt[3]{2}, \zeta\sqrt[3]{2}, \zeta^2\sqrt[3]{2}\}$ of $X^3 - 2$, is the full symmetric group.

The subgroups of G are G, $H_1 = \{1, \sigma, \sigma^2\}$, $H_2 = \{1, \tau\}$, $H_3 = \{1, \sigma\tau\}$, $H_4 = \{1, \sigma^2\tau\}$, and $\{1\}$. Its fixed fields $F = \mathbb{Q}$, F_1, F_2, F_3, F_4, and E satisfy $[F_1 : \mathbb{Q}] = 2$ and $[F_i : \mathbb{Q}] = 3$ for $i = 2, 3, 4$. Since $\zeta \in F_1$, $\sqrt[3]{2} \in F_2$, $\zeta^2\sqrt[3]{2} \in F_3$, and $\zeta\sqrt[3]{2} \in F_4$, we have $F_1 = \mathbb{Q}(\zeta)$, $F_2 = \mathbb{Q}(\sqrt[3]{2})$, $F_3 = \mathbb{Q}(\zeta^2\sqrt[3]{2})$, and $F_4 = \mathbb{Q}(\zeta\sqrt[3]{2})$. These are all the intermediate fields of $\mathbb{Q}(\sqrt[3]{2}, \zeta)/\mathbb{Q}$.

7 First Applications of Galois Theory

7.1 Definition A field F is called *algebraically closed*, if F/F is the only algebraic field extension of F.

Obviously, the following are equivalent:
(i) F is algebraically closed.
(ii) Every $f \in F[X] \smallsetminus F$ has a root in F.
(iii) Every $f \in F[X] \smallsetminus F$ is a product of linear factors in $F[X]$.
(iv) Every irreducible polynomial in $F[X]$ has degree 1.

7.2 Theorem (Fundamental theorem of algebra, Gauß 1777-1855) *$\mathbb{C}$ is algebraically closed.*

Proof Let $f \in \mathbb{C}[X] \smallsetminus \mathbb{C}$ and let $E/\mathbb{C}$ be a splitting field of f over $\mathbb{C}$. We have to show that $E = \mathbb{C}$. Since $\mathbb{C}/\mathbb{R}$ is finite, there exists a finite normal field extension $E'/\mathbb{R}$ by Corollary 4.4. Since $\mathrm{char}(\mathbb{R}) = 0$, $E'/\mathbb{R}$ is also separable and therefore a Galois extension. Let P be a Sylow 2-subgroup of $G := \mathrm{Gal}(E'/\mathbb{R})$ and set $E_2 := \mathrm{Fix}_{E'}(P)$. Then $[E_2 : \mathbb{R}] = [G : P]$ is odd. Since $E_2/\mathbb{R}$ is separable, there exists $a \in E_2$ with $E_2 = \mathbb{R}(a)$. Since $\deg(\min_{a/\mathbb{R}}) = [E_2 : \mathbb{R}]$ is odd, $\min_{a/\mathbb{R}}$ has a root in $\mathbb{R}$ by Analysis. Therefore, $\deg(\min_{a/\mathbb{R}}) = 1$, $E_2 = \mathbb{R}$, and $G = P$. This implies that $E'/\mathbb{C}$ is a Galois extension with $Q := \mathrm{Gal}(E'/\mathbb{C})$ of order 2^r for some $r \in \mathbb{N}_0$. Assume that $r \geqslant 1$. Then, by the Sylow theorems, Q has subgroup Q_1 of index 2, and Q_1 is normal in Q. Therefore $E_1 := \mathrm{Fix}_{E'}(Q_1)$ is a Galois extension of $\mathbb{C}$ of degree 2. Let $b \in E_1$ such that $\mathbb{C}(b) = E_1$ and let $\min_{b/\mathbb{C}} = X^2 + \alpha X + \beta$ with $\alpha, \beta \in \mathbb{C}$. But if $\gamma \in \mathbb{C}$ is a square root of $\alpha^2 - 4\beta$ (square roots exist in $\mathbb{C}$), then $(-\alpha \pm \gamma)/2 \in \mathbb{C}$ is a root of $\min_{b/\mathbb{C}}(X) = X^2 + \alpha X + \beta$. This is a contradiction. Therefore, $r = 0$, $E' = \mathbb{C}$, and $E = \mathbb{C}$. $\square$

7.3 Remark (Summary of the basic theory of finite fields) Let F be a finite field. Then $\mathrm{char}(F) = p$ for some prime p and F is a vector space over the subfield $P := \{0, 1, 1+1, \ldots, \}$ which has p elements and is isomorphic to $\mathbb{Z}/p\mathbb{Z}$. Therefore, $|F| = p^n$ for $n := [F : P]$. Moreover, $F^\times = \langle a \rangle$ for some $a \in F$ by Theorem 5.9. Obviously, every element in F is a root of $X^{p^n} - X \in P[X]$. Since $F = P(a)$, F is a splitting field of $X^{p^n} - X$ over P. Now Theorem 4.11 shows that any two fields with p^n elements are isomorphic. We know that F/P is separable, since P is perfect, and also normal, since F is a splitting field $X^{p^n} - X$. Therefore, F/P is a Galois extension. Let $\sigma \colon F \to F$, $x \mapsto x^p$ be the Frobenius automorphism. Then $\sigma \in G := \mathrm{Gal}(F/P)$ and for $k \in \mathbb{N}$ we have:

$$\sigma^k = \mathrm{id}_F \iff \sigma^k(a) = a \iff a^{p^k} = a\,.$$

Therefore, the order of σ is $n = [F : P] = G$ and $G = \langle \sigma \rangle$ is cyclic. By the main theorem of Galois theory, the subfields of F are given as $\mathrm{Fix}_F(\langle \sigma^m \rangle)$, with $m \mid n$. We have $[\mathrm{Fix}_F(\langle \sigma^m \rangle) : P] = [G : \langle \sigma^m \rangle] = m$. Therefore, $|\mathrm{Fix}_F(\langle \sigma^m \rangle)| = p^m$ and $\mathrm{Fix}_F(\langle \sigma^m \rangle) = P(a^{p^{n/m}})$.

We still do not know, if for every prime power p^n, with $n \in \mathbb{N}$, there exists a finite field with p^n elements. For that purpose, let E be a splitting field of $f(X) = X^{p^n} - X \in \mathbb{Z}/p\mathbb{Z}[X]$. Then certainly E is a finite field. Let again $\sigma \colon E \to E$ denote the Frobenius automorphism of E. Since $E = \mathbb{Z}/p\mathbb{Z}(R)$, where R is the set of roots of $X_X^{p^n}$ in E, and since $\sigma^n(r) = r$ for all $r \in R$ and $\sigma(\alpha) = \alpha$ for all $\alpha \in \mathbb{Z}/p\mathbb{Z}$, we have $\sigma^n(a) = a$ for all $a \in E$. Therefore, every element of E is a root of $f(X)$ and $|E| \leqslant p^n$. On the other hand, $f(X)$ has no multiple roots in E, since $f'(X) = -1$ and f are coprime (cf. Exercise 49). This shows that $|E| \geqslant p^n$. Hence, E is a field with p^n elements.

In order to construct a field with p^n elements it is usually easier to find a polynomial $f(X) \in \mathbb{Z}/p\mathbb{Z}[Z]$ of degree n which is irreducible. Then $\mathbb{Z}/p\mathbb{Z}[X]/(f(X))$ is a field with p^n elements.

7.4 Definition Let F be a field and let $n \in \mathbb{N}$.

(a) An element $\zeta \in F$ with ζ^n is called an *n-th root of unity*. If ζ has order n in $F^\times$, then ζ is called a *primitive* n-th root of unity. We write $\mu_n(F)$ for the multiplicative group of n-th roots of unity in F. Every $\zeta \in \mu_n(F)$ is a primitive d-th root of unity for a unique divisor d of n. Conversely $\mu_d(F) \leqslant \mu_n(F)$ for each divisor d of n. Therefore, if we write $\tilde{\mu}_n(F) \subseteq \mu_n(F)$ for the subset of primitive n-th roots of unity in F, the group $\mu_n(F)$ is the disjoint union of the subsets $\tilde{\mu}_d(F)$, where d runs through all positive divisors of n.

(b) Every splitting field of $X^n - 1$ over F is called the *n-th cyclotomic field over F* and is denoted by F_n. The polynomial

$$\Phi_n(X) := \prod_{\zeta \in \tilde{\mu}_n(F_n)} (X - \zeta) \in F_n[X]$$

is called the *n-th cyclotomic polynomial over F*.

7.5 Remark Let F be a field and let $n \in \mathbb{N}$.

(a) Since $X^n - 1$ can have at most n roots in F, we have $|\mu_n(F)| \leqslant n$, and $\mu_n(F)$ is cyclic by Theorem 5.9.

(b) If F contains a primitive n-th root of unity ζ, then $|\langle\zeta\rangle| = n$ and therefore, $\mu_n(F) = \langle\zeta\rangle$. For $i \in \mathbb{Z}$ we then have: ζ^i is again a primitive n-th root of unity if and only if $\gcd(i,n) = 1$. Therefore, $|\tilde{\mu}_n(F)|$ equals the number of elements $i + n\mathbb{Z} \in \mathbb{Z}/n\mathbb{Z}$ with $\gcd(i,n) = 1$, which are exactly the units in $\mathbb{Z}/n\mathbb{Z}$. Their number is $\varphi(n)$, where φ denotes the Euler φ-function. In particular, for the n-th cyclotomic polynomial $\Phi_n \in F_n[X]$ we have either $\Phi_n = 1$ or Φ_n has degree $\varphi(n)$, depending on whether there exists a primitive n-th root of unity in F_n or not.

(c) If $\mathrm{char}(F) = p > 0$ and $n = pm$, then there are no primitive n-th roots of unity in F, since $0 = \zeta^{pm} - 1 = (\zeta^m - 1)^p$ for every $\zeta \in \mu_n(F)$, and so ζ has order at most m. In particular, this holds for F_n instead of F, so that $\Phi_n(X) = 1$ in this case.

(d) If $\mathrm{char}(F) \nmid n$ (these are the remaining cases, i.e., $\mathrm{char}(F) = 0$ or $\mathrm{char}(F) = p > 0$ with $p \nmid n$), then $(X^n - 1)' = nX^{n-1}$ is coprime to $X^n - 1$. Therefore, by Exercise 49, the roots of $X^n - 1$ in F_n are pairwise distinct, and they form a cyclic group $\langle\zeta\rangle$ of order n. This implies that F_n contains a primitive n-th root of unity ζ. Moreover, $F_n = F(\zeta)$, since $X^n - 1 = \prod_{i=1}^n (X - \zeta^i)$.

(e) Assume that $\mathrm{char}(F) \nmid n$. Then $\deg(\Phi_n(X)) = \varphi(n)$, and since $\mu_n(F_n)$ is the disjoint union of $\tilde{\mu}_d(F_n)$, where $d \mid n$, we obtain

$$X^n - 1 = \prod_{\zeta \in \mu_n(F_n)} (X - \zeta) = \prod_{d \mid n} \Phi_d(X)\,.$$

In particular, $\Phi_n(X)$ can be determined recursively by

$$\Phi_n(X) = \frac{X^n - 1}{\prod_{d|n,\ d\neq n} \Phi_d(X)}\,.$$

Note that since each $\Phi_d(X)$ has leading coefficient 1, and since $\Phi_1(X) = X - 1$ has coefficients in $\mathbb{Z}\cdot 1_F$, also Φ_n has coefficients in $\mathbb{Z}\cdot 1_F$ by an inductive argument. If F has characteristic 0, the coefficients lie in $\mathbb{Z}$, and if F has prime characteristic p, then the coefficients lie in the unique subfield of F of order p. In particular, if we compute $\Phi_n(X)\in\mathbb{Z}[X]$ for $F=\mathbb{Q}$, then $\Phi_n\in\mathbb{Z}[X]$ and we obtain $\Phi_n(X)$ for an arbitrary field F with $\mathrm{char}(F)\nmid n$ by using the unique ring homomorphism $\mathbb{Z}\to F$ and extending it to a ring homomorphism $\mathbb{Z}[X]\to F[X]$ in the obvious way. In the sequel we will usually mean the cyclotomic polynomial with integer coefficients when we use the symbol $\Phi_n(X)$.

7.6 Example (a) $\mu_n(\mathbb{R}) = \{1,-1\}$ for every even n.

(b) $\zeta := e^{2\pi i/n}\in\mathbb{C}$ is a primitive n-th root of unity in $\mathbb{C}$.

(c) If F is a finite field with p^n elements, then $\mu_{p^n-1}(F) = F^\times$ and F contains a primitive p^n-1-th root of unity.

(d) If p is a prime, then

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

7.7 Theorem *Let F be a field and let $n\in\mathbb{N}$ with $\mathrm{char}(F)\nmid n$. Then F_n/F is a Galois extension with $G := \mathrm{Gal}(F_n/F)$ isomorphic to a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. In particular, G is abelian.*

Proof By Remark 7.5(d), X^n-1 is separable over F. Therefore, F_n/F is a Galois extension. Also, by Remark 7.5(d), $F_n = F(\zeta)$ with a primitive n-th root of unity ζ. Clearly, $\sigma(\zeta)$ is again a primitive n-th root of unity for all $\sigma\in G$. Therefore, $\sigma(\zeta)=\zeta^i$ for a unique $i+n\mathbb{Z}\in(\mathbb{Z}/n\mathbb{Z})^\times$. Moreover, if $\tau\in G$ and $\tau(\zeta)=\zeta^j$ with $j+n\mathbb{Z}\in(\mathbb{Z}/n\mathbb{Z})^\times$, then $(\sigma\tau)(\zeta)=\sigma(\zeta^j)=\sigma(\zeta)^j=\zeta^{ij}$. Therefore, the resulting map $G\to(\mathbb{Z}/n\mathbb{Z})^\times$ is a group homomorphism. It is injective, since $\sigma\in G$ is uniquely determined by $\sigma(\zeta)$. $\square$

7.8 Remark Let F and $n\in\mathbb{N}$ be as in Theorem 7.7. Since $G:=\mathrm{Gal}(F_n/F)$ is abelian, every subgroup of G is normal in G. Therefore, every intermediate field E in F_n/F is a Galois extension of F and its Galois group is a factor group of G and hence abelian. A kind of converse for the base field $\mathbb{Q}$ to that was proved by Kronnecker and Weber using techniques from algebraic number theory: Let $F\subset\mathbb{C}$, then $F/\mathbb{Q}$ is a Galois extension with abelian Galois group if and only if $F\subseteq\mathbb{Q}(\zeta)$ for some root of unity $\zeta\in\mathbb{C}$.

7.9 Theorem *For each $n\in\mathbb{N}$, $\Phi_n(X)\in\mathbb{Z}[X]$ is irreducible in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$. Moreover, if $\zeta\in\mathbb{C}$ is a primitive n-th root of unity, then $\mathbb{Q}_n:=\mathbb{Q}(\zeta)$ is a Galois extension of $\mathbb{Q}$ of degree $\varphi(n)$ and $(\mathbb{Z}/n\mathbb{Z})^\times\to\mathrm{Gal}(\mathbb{Q}_n(\zeta)/\mathbb{Q})$, $i+n\mathbb{Z}\mapsto\sigma_i$, with $\sigma_i(\zeta)=\zeta^i$, is a group isomorphism.*

Proof It suffices to show that $\Phi_n(X)$ is irreducible in $\mathbb{Z}[X]$. The rest follows from Gauß' Lemma and from Theorem 7.7 and its proof.

Assume that $f \in \mathbb{Z}[X]$ is an irreducible monic factor of Φ_n, and let $X^n - 1 = f(X)g(X)$ with $g(X) \in \mathbb{Q}[X]$. Then $g(X) \in \mathbb{Z}[X]$, since $f(X)$ is monic. Let $\eta \in \mathbb{Q}_n$ be a root of f. Then $f = \min_{\eta/\mathbb{Q}}$ and $\eta \in \tilde{\mu}_n(\mathbb{Q}_n)$.

Claim: For every prime p with $p \nmid n$ we have $f(\eta^p) = 0$. The claim is proved as follows. Assume that $f(\eta^p) \neq 0$ for some prime $p \nmid n$. Then $f(\eta^p)g(\eta^p) = (\eta^p)^n - 1 = 1 - 1 = 0$ and $g(\eta^p) = 0$. Since $f = \min_{\eta/\mathbb{Q}}$, there exists $h \in \mathbb{Q}[X]$ such that $g(X^p) = f(X)h(X)$. Since f is monic, $g \in \mathbb{Z}[X]$ implies $h \in \mathbb{Z}[X]$. Let $\varphi\colon \mathbb{Z}[X] \to \mathbb{Z}/p\mathbb{Z}[X]$ be the ring homomorphism with $\varphi(a_nX^n + \cdots + a_0) = \bar{a}_nX^n + \cdots + \bar{a}_0$. Since f is monic, $\varphi(f)$ is not constant and has an irreducible factor $q \in \mathbb{Z}/p\mathbb{Z}[X]$. Write $g = X^k + a_{k-1}X^{k-1} + \cdots + a_0$ with $a_{k-1}, \ldots, a_0 \in \mathbb{Z}$. Since $\bar{a}^p = \bar{a}$ for all $a \in \mathbb{Z}$, we have

$$\begin{aligned} \varphi(f)\varphi(h) = \varphi(g(X^p)) &= X^{kp} + \bar{a}_{k-1}X^{(k-1)p} + \cdots + \bar{a}_0 \\ &= (X^k + \bar{a}_{k-1}X^{k-1} + \cdots + \bar{a}_0)^p = \varphi(g)^p \,. \end{aligned}$$

Since $q \mid \varphi(f)$, also $q \mid \varphi(g)$ and $q^2 \mid \varphi(f \cdot g) = \varphi(X^n - 1) = X^n - \bar{1}$. This contradicts Remark 7.5(d), and the claim is proved.

Now, η^p is a root of f for all primes p with $p \nmid n$. By induction on r we obtain that $\eta^{p_1 \cdots p_r}$ is a root of f for arbitrary (not necessarily distinct) primes $p_1, \ldots, p_r$ not dividing n. Therefore, $f(\eta^m) = 0$ for all $m \in \mathbb{N}$ with $\gcd(m, n) = 1$, which shows that each primitive n-th root of unity is a root of f. This implies $f = \Phi_n$. $\square$

7.10 Theorem *Let $n = 2^{\alpha} p_1^{\alpha_1} \cdots p_t^{\alpha_t} \in \mathbb{N}$ with $\alpha \in \mathbb{N}_0$, $\alpha_1, \ldots, \alpha_t \in \mathbb{N}$, and with primes $3 \leqslant p_1 < p_2 < \cdots < p_t$. The regular n-gon is constructible from 0 and 1 if and only if $p_1, \ldots, p_t$ are* Fermat primes *(these are primes p of the form $p = 2^{2^r} + 1$ for some $r \in \mathbb{N}_0$) and $\alpha_1 = \cdots = \alpha_t = 1$.*

Proof The regular n-gon is constructible if and only if $\zeta := e^{2\pi i/n}$ is constructible. By Theorem 3.4, this is equivalent to $\mathbb{Q}(\zeta)$ being contained in the top field of a square root tower $\mathbb{Q} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_m$. This in turn is equivalent to $[\mathbb{Q}(\zeta) : \mathbb{Q}]$ being a power of 2. But

$$[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(n) = 2^{\alpha-1}(p_1 - 1) \cdots (p_t - 1) p_1^{\alpha_1 - 1} \cdots p_t^{\alpha_t - 1} \,,$$

and we obtain the equivalent condition $\alpha_1 = \cdots \alpha_t = 1$ and $p_i = 2^{\beta_i} + 1$ with some $\beta_i \in \mathbb{N}$, for all $i = 1, \ldots, t$. But note, that if $2^s + 1$ is a prime, then necessarily s is a power of 2, because otherwise $s = qt$ with an odd prime q and $t \in \mathbb{N}$ and we obtain a decomposition $2^s + 1 = (2^t + 1)((2^t)^{q-1} - (2^t)^{q-2} + - \cdots + 1)$. This completes the proof. $\square$

8 Solvability of Polynomial Equations by Radicals

8.1 Definition A group G is called *solvable*, if there exist finitely many subgroups

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = 1$$

such that G_{i-1}/G_i is abelian for all $i = 1, \ldots, n$.

8.2 Lemma *Let G be a group, let $H \leqslant G$ and $N \trianglelefteq G$.*
(a) If G is solvable, then also G and G/N are solvable.
(b) G is solvable if and only if G/N and N are solvable.

Proof (a) Let $G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = 1$ with G_{i-1}/G_i abelian for all $i = 1, \ldots, n$. Then $H = H \cap G_0 \geqslant H \cap G_1 \geqslant \cdots \geqslant H \cap G_n = 1$ with $H \cap G_i \trianglelefteq H \cap G_{i-1}$ and

$$\begin{aligned} H \cap G_{i-1}/H \cap G_i &= H \cap G_{i-1}/(H \cap G_{i-1}) \cap G_i \\ &\cong (H \cap G_{i-1})G_i/G_i \leqslant G_{i-1}/G_i \end{aligned}$$

abelian. This shows that H is solvable. Moreover,

$$G/N = G_0N/N \geqslant G_1N/N \geqslant \cdots \geqslant G_nN/N = N/N$$

with $G_iN/N \trianglelefteq G_{i-1}N/N$ and

$$\begin{aligned} (G_{i-1}N/N)/(G_iN/N) &\cong G_{i-1}N/G_iN = G_{i-1}(G_iN)/G_iN \\ &\cong G_{i-1}/G_{i-1} \cap G_iN \cong (G_{i-1}/G_i)/(G_{i-1} \cap G_iN/G_i) \end{aligned}$$

abelian. Therefore, G/N is solvable.

(b) By (a) we only have to show that if N and G/N are solvable, then also G is solvable. Let

$$G/N = X_0 \trianglerighteq X_1 \trianglerighteq \cdots \trianglerighteq X_n = N/N$$

and

$$N = N_0 \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_m = 1$$

such that X_{i-1}/X_i and N_{j-1}/N_j is abelian for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Then $X_i = G_i/N$ for a unique subgroup G_i with $N \leqslant G_i \leqslant G$ for $i = 1, \ldots, n$. Moreover, since $G_i/N \trianglelefteq G_{i-1}/N$, also $G_i \trianglelefteq G_{i-1}$ and $G_{i-1}/G_i \cong (G_{i-1}/N)/(G_i/N)$ is abelian. Therefore, the chain

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_0 = N = N_0 \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_m = 1$$

has the right properties to imply that G is solvable. □

8.3 Example The symmetric groups S_1, S_2, S_3, S_4 are solvable. In fact, S_1 and S_2 are abelian, and $S_3 \trianglerighteq A_3 \trianglerighteq 1$ and $S_4 \trianglerighteq A_4 \trianglerighteq V_V \trianglerighteq 1$ with $V_4 := \{1, (12)(34), (13)(24), (14)(23)\}$ and abelian factor groups S_3/A_3, $A_3/1$, S_4/A_4, A_4/V_4, $V_4/1$.

8.4 Theorem *S_n is solvable if and only if $n \leqslant 4$.*

Proof By Example 8.3, we only have to show that S_n is not solvable for $n \geqslant 5$. By Lemma 8.2, it suffices to show that the subgroup $H \leqslant S_n$ which is generated by the set of 3-cycles (i, j, k), is not solvable. Assume that H is solvable. Then there exists $H_1 \triangleleft H$ such that H/H_1 is abelian. In particular $H' := \langle aba^{-1}b^{-1} \mid a, b \in H\rangle \leqslant H_1 < H$. But for each 3-cycle (i, j, k) there exist elements $l, m \in \{1, \ldots, n\}$ such that $|\{i, j, k, l, m\}| = 5$ and we obtain

$$(i, j, l)(i, k, m)(i, j, l)^{-1}(i, k, m)^{-1} = (i, j, k)$$

implying that $H' = H$ and contradicting $H' < H$. Therefore, H and S_n are not solvable. □

8.5 Theorem *Let F be a field and let $n \in \mathbb{N}$ be such that $\mathrm{char}(F) \nmid n$. Moreover, let $\alpha \in F$ and let E be a splitting field of $X^n - \alpha$ over F.*

(a) E contains a primitive $n-th$ root of unity ζ.

(b) If $b \in E$ is such that $b^n = \alpha$, then $b, \zeta b, \ldots, \zeta^{n-1}b$ are the roots of $X^n - \alpha$. In particular, $X^n - \alpha$ has n pairwise distinct roots and $E = F(b, \zeta)$.

(c) E/F, $E/F(\zeta)$, and $F(\zeta)/F$ are Galois extensions.

(d) $\mathrm{Gal}(E/F(\zeta))$ is isomorphic to a subgroup of $\mathbb{Z}/n\mathbb{Z}$. In particular, if F contains a primitive n-th root of unity, then E/F has a cyclic Galois group.

Proof (a) $(X^n - \alpha)' = nX^{n-1}$ and $X^n - \alpha$ are coprime. Therefore, $X^n - \alpha$ has n pairwise distinct roots $b_1, \ldots, b_n$ in E by Exercise 49. Then $b_1/b_1, b_2/b_1, \ldots, b_n/b_1$ are pairwise distinct n-th roots of unity, since $(b_i/b_1)^n = \alpha/\alpha = 1$. Therefore, $|\mu_n(E)| = n$. Moreover, $\mu_n(E)$ is cyclic by Theorem 5.9, and hence generated by a primitive n-th root of unity ζ.

(b) This is obvious.

(c) E/F is a Galois extension, since E is the splitting field of the separable polynomial $X^n - \alpha$. $F(\zeta)/F$ is a Galois extension by Theorem 7.7. $E/F(\zeta)$ is a Galois extension, since E/F is.

(d) Let $\sigma \in \mathrm{Gal}(E/F(\zeta))$. Then $\sigma(b) \in \{b, \zeta b, \zeta^2 b, \ldots, \zeta^{n-1}b\}$ and $\sigma(b)/b \in \langle\zeta\rangle$. This defines a function $\varphi\colon \mathrm{Gal}(E/F(\zeta)) \to \langle\zeta\rangle$. For $\sigma, \tau \in \mathrm{Gal}(E/F(\zeta))$ we have:

$$\varphi(\sigma\tau) = \frac{\sigma(\tau(b))}{b} = \frac{\sigma(b\frac{\tau(b)}{b})}{b} = \frac{\sigma(b)\frac{\tau(b)}{b}}{b} = \varphi(\sigma)\varphi(\tau)\,,$$

since $\tau(b)/b \in F(\zeta)$. Therefore, φ is a group homomorphism. If $\varphi(\sigma) = 1$, then $\sigma(b) = b$, hence $\sigma = \mathrm{id}_E$, since $E = (F(\zeta))(b)$. □

8.6 Theorem *Let $n \in \mathbb{N}$ and assume that F is a field containing a primitive n-th root of unity ζ. If E/F is a Galois extension with cyclic Galois group $G = \langle \sigma \rangle$ of order n, then there exists $\alpha \in F$ such that E is a splitting field of $X^n - \alpha$ over F.*

Proof By Lemma 6.3, there exists $c \in E$ such that

$$b := c + \zeta\sigma(c) + \zeta^2\sigma^2(c) + \cdots + \zeta^{n-1}\sigma^{n-1}(c) \neq 0 .$$

We have

$$\sigma(b) = \sigma(c) + \zeta\sigma^2(c) + \cdots + \zeta^{n-2}\sigma^{n-1}(c) + \zeta^{n-1}b = \zeta^{-1}b ,$$

and therefore, $\sigma(F(b) = F(b)$ and $\sigma^m(b) = \zeta^{-m}b$ for all $m \in \mathbb{N}$. Therefore, the restrictions of $1, \sigma, \sigma^2, \ldots, \sigma^{n-1}$ form n pairwise distinct elements in $\mathrm{Aut}(F(b)/F)$ and we obtain

$$n \leqslant |\mathrm{Aut}(F(b)/F)| \leqslant [F(b) : F] = [E : F] = n ,$$

which implies equality everywhere and $F(b) = E$. Since $\sigma(b^n) = \sigma(b)^n = (\zeta b)^n = b^n$, we have $\alpha := b^n \in F$. Obviously, E is a splitting field of $X^n - \alpha$ over F, since $b, \zeta b, \ldots, \zeta^{n-1}b$ are the roots of $X^n - \alpha$. $\square$

8.7 Definition A field extension E/F is called a *radical extension* if there exists a chain $F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_r = E$ and elements $b_i \in F_i$, $n_i \in \mathbb{N}$ for $i = 1, \ldots, r$ such that $F_i = F_{i-1}(b_i)$ and $b_i^{n_i} \in F_{i-1}$ for all $i = 1, \ldots, r$. In particular, $[E : F]$ is finite. A polynomial equation $f(X) = 0$ with irreducible $f \in F[X]$ is called *solvable by radicals over F*, if there exists a radical extension E/F such that $f(X)$ splits into linear factors in $E[X]$.

8.8 Lemma *Let F be a field of characteristic 0 and let E/F be a radical extension. Then there exists an extension E'/E such that E'/F is a radical Galois extension.*

Proof Let $F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_r = F$, $b_i \in F_i$, $n_i \in \mathbb{N}$ be as in Definition 8.7. By Corollary 4.14, there exists a finite field extension $\tilde{E}/E$ such that $\tilde{E}/F$ is normal and hence a Galois extension. We write $G := \mathrm{Gal}(\tilde{E}/F) = \{\sigma_1 = \mathrm{id}_{\tilde{E}}, \sigma_2, \ldots, \sigma_s\}$ and set

$$F_{ij} := F(\sigma_1(b_1), \ldots, \sigma_s(b_1), \sigma_1(b_2) \ldots, \sigma_s(b_2), \ldots \sigma_1(b_i), \ldots, \sigma_j(b_i))$$

for $i = 1, \ldots, r$ and $j = 1, \ldots, s$. Then we have a tower of fields

$$F \subseteq F_{11} \subseteq F_{12} \subseteq \cdots \subseteq F_{1s} \subseteq F_{21} \subseteq \cdots \subseteq F_{2s} \subseteq \cdots \subseteq F_{rs} \subseteq \tilde{E} .$$

For each $i = 1, \ldots, r$, F_{is}/F is a Galois extension. In fact, each $\tau \in G$ permutes the elements $\sigma_1(b_1), \ldots, \sigma_s(b_1), \ldots, \sigma_1(b_i), \ldots, \sigma_s(b_i)$, and therefore, $\tau(F_{is}) \subseteq$

F_{is} and Theorem 6.13 applies. In particular, $E' := F_{rs}$ is a Galois extension of F.

Obviously, $E = F(b_1, \ldots, b_r) \subseteq E'$. Moreover, for $i = 1, \ldots, r$ and $j = 2, \ldots, s$, we have

$$\sigma_j(b_i)^{n_i} = \sigma_j(b_i^{n_i}) \in \sigma_j(F_{i-1}) \subseteq \sigma_j(F_{i-1,s}) = F_{i-1,s} \subseteq F_{i,j-1}\,.$$

Moreover, $\sigma_1(b_1)^{n_1} = b_1^{n_1} \in F$ and for $i = 2, \ldots, n$ we have

$$\sigma_1(b_i)^{n_i} = b_i^{n_i} \in F_{i-1} \subseteq F_{i-1,1} \subseteq F_{i-1,s}\,.$$

Therefore, E'/F is also a radical extension. □

8.9 Corollary *Let F be a field of characteristic 0 and let $f(X) \in F[X]$ be irreducible. Assume that f has a root in a radical field extension E/F. Then there exists a radical field extension E'/F such that $f(X)$ splits completely into linear factors in $E'[X]$.*

Proof By Lemma 8.8 there exists a radical Galois extension E'/F with $E \subseteq E'$. Since E'/F is normal and contains a root of f, f splits completely into linear factors in $E'[X]$. □

8.10 Theorem *Let F be a field of characteristic 0 and let E/F be a radical Galois extension. Then $G := \mathrm{Gal}(E/F)$ is solvable.*

Proof Let $F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_r = E$, $b_i \in F_i$, $n_i \in \mathbb{N}$ be as in Definition 8.7. Set $n = n_1 \cdots n_r$ and let $\tilde{E} := E(\zeta)$ with a primitive n-th root of unity ζ. Since E/F and $F(\zeta)/F$ are Galois extensions, also $\tilde{E} \cdot F(\zeta)/F$ is a Galois extension, because $E \cdot F(\zeta)/F$ is normal by Exercise 37. By Lemma 8.2, it suffices to show that $\tilde{G} := \mathrm{Gal}(\tilde{E}/F)$ is solvable. The chain

$$F = F_0 \subseteq F_0(\zeta) \subseteq F_1(\zeta) \subseteq \cdots \subseteq F_r(\zeta) = E(\zeta) = \tilde{E}$$

corresponds to a chain of subgroups

$$\tilde{G} \geqslant \tilde{G}_0 \geqslant \tilde{G}_1 \geqslant \cdots \geqslant \tilde{G}_r = 1\,.$$

Since $F(\zeta)/F$ is a Galois extension, $\tilde{G}_0 \trianglelefteq \tilde{G}$ and $\tilde{G}/\tilde{G}_0 = \mathrm{Gal}(F(\zeta)/F)$ is abelian by Theorem 7.7. Moreover, for $i = 1, \ldots, r$, we have $F_i(\zeta) = F_{i-1}(b_i, \zeta) = (F_{i-1}(\zeta))(b_i)$ with $b_i^{n_i} \in F_{i-1}(\zeta)$. Therefore, $F_i(\zeta)$ is a splitting field of $X^{n_i} - b_i^{n_i}$ over $F_{i-1}(\zeta)$. By Theorem 8.5(c) and (d), $F_i(\zeta)/F_{i-1}(\zeta)$ is a Galois extension with cyclic Galois group. Therefore, $\tilde{G}_i \trianglelefteq \tilde{G}_{i-1}$ and $G_{i-1}/G_i = \mathrm{Gal}(F_i(\zeta)/F_{i-1}(\zeta))$ is abelian. This implies that $\tilde{G}$ is solvable. □

8.11 Theorem *Let E/F be a Galois extension with solvable Galois group G of order n and assume that $\mathrm{char}(F) = 0$. Then the n-th cyclotomic field $E(\zeta)$, where ζ a primitive n-th root of unity, is a radical Galois extension of F.*

Proof Since $F(\zeta)/F$ and E/F are Galois extensions, also $E(\zeta) = E \cdot F(\zeta)/F$ is a Galois extension. By Exercise 52, $E(\zeta)/F(\zeta)$ is a Galois extension with

$$G := \mathrm{Gal}(E(\zeta)/F(\zeta)) \cong \mathrm{Gal}(E/E \cap F(\zeta)) \leqslant \mathrm{Gal}(E/F)\,.$$

Therefore, also G is solvable and $|G| \mid n$. Let

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_r = 1$$

be as in Definition 8.1. Since every finite abelian group is a direct sum of cyclic groups, we may even assume that G_{i-1}/G_i is cyclic for $i = 1, \ldots, r$, by refining the above chain if necessary. Consider the intermediate fields $F_i := \mathrm{Fix}_{E(\zeta)}(G_i)$ in $E(\zeta)/F(\zeta)$ for $i = 1, \ldots, r$. They form a chain

$$F \subseteq F(\zeta) = F_0 \subseteq \cdots \subseteq F_r = E(\zeta)\,,$$

with F_i/F_{i-1} a Galois extension, since $G_i \trianglelefteq G_{i-1}$ for all $i = 1, \ldots, r$. Moreover, $\mathrm{Gal}(F_i/F_{i-1}) \cong G_{i-1}/G_i$ is cyclic. By Theorem 8.6, there exists $b_i \in F_i$ with $b_i^{n_i} \in F_{i-1}$. Moreover, $\zeta^n \in F$. Therefore, $E(\zeta)/F$ is a radical extension. □

8.12 Theorem *Let F be a field of characteristic 0, let $f \in F[X]$ be irreducible, let E be a splitting field of f over F, and set $G := \mathrm{Gal}(E/F)$, the Galois group of f over F. Then the following are equivalent:*

(i) *$f(X) = 0$ is solvable by radicals over F.*

(ii) *G is solvable.*

Proof (i)$\Rightarrow$(ii): Let E'/F be a radical extension such that $f(X)$ splits into linear factors in $E'[X]$. By Lemma 8.8 we may choose E' such that E'/F is also a Galois extension. By Theorem 8.10, $\mathrm{Gal}(E'/F)$ is solvable. Let $r_1, \ldots, r_n \in E'$ be the roots of f. Then $F(r_1, \ldots, r_n) \subseteq E'$ is a splitting field of f over F, hence G is isomorphic to the factor group $\mathrm{Gal}(F(r_1, \ldots, r_n)/F)$ of the solvable group $\mathrm{Gal}(E'/F)$. Hence G is solvable by Lemma 8.2.

(ii)$\Rightarrow$(i): If G is solvable, there exists a radical extension E'/F with $E \subseteq E'$ by Theorem 8.11. Since f splits into linear factors in $E[X]$, it also does in $E'[X]$. □

8.13 Lemma *Let p be a prime.*

(a) *Each element $\sigma \in S_p$ of order p is a p-cycle.*

(b) *If $\sigma \in S_p$ has order p and $\tau \in S_p$ is a transposition, then $\langle \sigma, \tau \rangle = S_p$.*

Proof (a) Assume that σ is not a p-cycle. Then, in the cycle notation it contains a k-cycle $(a_1, \ldots, a_k)$ for some $k \in \{2, \ldots, p-1\}$. This implies that $\sigma^k(a_1) = a_1 = \sigma^p(a_1)$. Since k and p are coprime, it also implies $\sigma(a_1) = a_1$ in contradiction to $\sigma(a_1) = a_2$.

(b) Since σ has order p, σ is a p-cycle by (a). Write $\sigma = (a_1, \ldots, a_p)$ and $\tau = (a_i, a_j)$. Since also $\sigma = (a_i, \ldots, a_p, a_1, \ldots, a_{i-1})$, we can write $\sigma = (b_1, \ldots, b_p)$ and $\tau = (b_1, b_k)$ with $k \in \{2, \ldots, p\}$. Then $\sigma^{k-1} = (b_1, b_k, \ldots)$ is again a p-cycle by (a), and we can write $\sigma^{k-1} = (c_1, c_2, \ldots, c_p)$ and $\tau = (c_1, c_2)$ for certain $c_1, \ldots, c_p$. Define $\rho \in S_p$ by $\rho(i) := c_i$ for $i = 1, \ldots, p$. Then $\rho^{-1}\sigma^{k-1}\rho = (1, 2, 3, \ldots, p)$ and $\rho^{-1}\tau\rho = (1, 2)$. It is well-known that $S_p = \langle (1, 2, \ldots, p), (1, 2) \rangle$ and we obtain

$$\begin{aligned} S_p = \rho\langle (1, 2, \ldots, p), (1, 2)\rangle\rho^{-1} &= \langle \rho(1, 2, \ldots, p)\rho^{-1}, \rho(1, 2)\rho^{-1}\rangle \\ &= \langle \sigma^{k-1}, \tau\rangle \leqslant \langle \sigma, \tau\rangle \,. \end{aligned}$$

□

8.14 Theorem *Let p be a prime and let $f(X) \in \mathbb{Q}[X]$ be an irreducible polynomial of degree p which has exactly $p-2$ real roots. Then the Galois group of f is isomorphic to S_p. In particular, $f(X) = 0$ is not solvable by radicals over $\mathbb{Q}$, if $p \geqslant 5$.*

Proof Let $f = c(X - a_1) \cdots (X - a_p)$ be a decomposition in $\mathbb{C}[X]$. Then $E := \mathbb{Q}(a_1, \ldots, a_p)$ is a splitting field of f over $\mathbb{Q}$. $E/\mathbb{Q}$ is a Galois extension and we denote its Galois group by G. For each $i = 1, \ldots, p$ also $\bar{a}_i$ is a root of f. Therefore, the restriction of the complex conjugation is an element $\tau \in G$, which fixes the $p-2$ real roots and interchanges the two non-real roots. On the other hand, since E contains $\mathbb{Q}(a_1)$ and since $[\mathbb{Q}(a_1) : \mathbb{Q}] = \deg(\min_{a_1/\mathbb{Q}}) = \deg(f) = p$, also $p \mid |G|$ and G contains an element σ of order p. The action of G on $\{a_1, \ldots, a_p\}$ gives rise to a monomorphism $\varphi\colon G \to S_p$ as in Definition 6.14. Then $\varphi(\tau)$ is a transposition and $\varphi(\sigma)$ is an element of order p. By the previous lemma, the homomorphism φ is also surjective. □

8.15 Example The polynomial $f(X) = X^5 - 4X + 2$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$ by Eisenstein. Since $f'(X) = 5X^4 - 4$ has exactly two real roots, namely $\pm\sqrt[4]{4/5}$, and since $f(\sqrt[4]{4/5}) < 0$ and $f(-\sqrt[4]{4/5}) > 0$, f has exactly 3 real roots. Therefore, the Galois group of f is isomorphic to S_5, and its roots lie in no radical extension E/F.

References

[J] Jacobson: Basic Algebra I, 2nd edition, Freeman 1985.