Chapter 1 : Algebra preliminaries

Algebraic number theory, as the name suggests, employs a fair amount of results from abstract algebra far beyond the scope of one or two introductory courses on algebra. In this chapter I highlight some of the more advanced topics in algebra, that will prove useful during the rest of the course. The reader may, however, start wondering: "Where are my numbers?" True! We will eventually study *numbers* in latter chapters. This chapter is merely a collection of auxiliary results. I could have taken the approach: "Discuss algebra whenever needed". But this compilation of algebraic techniques is useful elsewhere in mathematics too. Having all these results at one place is expected to aid the reader not only by making referencing easy later but also by decoupling algebra from number theory to emphasize the independence (and yet the interdependence) of these two branches of mathematics.

Fortunately enough, throughout this course (as well as in the most part of the study of algebraic number theory) we don't have to dirty our hands by attempting to deal with non-commutativity. In fact, very little is known in non-commutative algebra which turns out to be a very difficult branch of mathematics. But here we start with the following glorious assumption:

1.1 Assumption By a ring we will always mean a *commutative ring with identity*.

Many results discussed in this chapter are valid for non-commutative rings too. But we won't even bother classifying the results based on their generalities. We will mostly study a special class of subrings of $\mathbb{C}$, the field of complex numbers, and all these rings naturally inherit commutativity from $\mathbb{C}$.

1.1 Divisibility in rings

The concept of divisibility naturally extends from $\mathbb{Z}$ to a general *ring* (i.e., a commutative ring with identity). However, one may not have unique factorization in all rings. $\mathbb{Z}$ is, indeed, a very well-mannered algebraic structure. Or perhaps I would better say that our notions of algebraic etiquette stem from studies of $\mathbb{Z}$ and other instances of rings that occurred 'naturally' to our mathematical ancestors.

1.2 Definition Let A be a ring.

- An element $a \in A$ is said to divide an element $b \in A$, if there exists $c \in A$ such that $b = ac$. In this case, we write $a \mid b$. If no such c exists, we write $a \not| b$.
- An element $u \in A$ is called a unit (of A), if $u \mid 1$, i.e., if there exists $v \in A$ with $uv = 1$. All the units of A form a group under the multiplication of A. This group is called the group of units of A and is denoted by A^*. A is a field, if and only if $A \neq \{0\}$ and $A^* = A \setminus \{0\}$. If $u \in A$ is a unit and $uv = 1$, we often write $v = u^{-1}$ and $u = v^{-1}$ and say that v (resp. u) is the inverse of u (resp. v).
- Two non-zero elements $a, b \in A$ are called associates, denoted by $a \sim b$, if $a = ub$ for some unit u of A. Clearly, $\sim$ is an equivalence relation on $A \setminus \{0\}$.
- A non-zero non-unit $p \in A$ is called a prime, if whenever $p \mid ab$ (with $a, b \in A$), we have either $p \mid a$ or $p \mid b$.
- A non-zero non-unit a is called irreducible, if any factorization $a = bc$ (with $b, c \in A$) implies that either b or c is a unit.

1.3 Example All the concepts introduced in Definition 1.2 pertain to the underlying ring A. If $A = \mathbb{Z}$, the only units we have are ± 1 and, therefore, $a \sim b$, if and only if $a = \pm b$. The primes are the *prime numbers* $2, 3, 5, 7, \ldots$ and the negatives of them.

If, on the other hand, we take $A = \mathbb{Q}$, every nonzero element of A becomes a unit and hence associate to one another. The 'integers' $\pm 2/1, \pm 3/1, \pm 5/1, \ldots$ are, in particular, units too and hence are neither prime nor irreducible. In fact, $\mathbb{Q}$ (or any field) contains no primes or irreducible elements. Also note that $2 \not| 3$ in $\mathbb{Z}$, but $2 \mid 3$ in $\mathbb{Q}$, since $3 = (3/2) \times 2$ and $3/2$ is a rational number but not an integer.

An integer p is prime (in $\mathbb{Z}$) if and only if p is irreducible (in $\mathbb{Z}$). This is, in fact, true for more general classes of rings as we will see below. We can now prove a partial result in this direction.

1.4 Proposition Let A be an integral domain and p a prime in A. Then p is irreducible.

Proof Let us write $p = ab$. We will show that either a or b is a unit. Clearly, $p \mid ab$ and hence $p \mid a$ or $p \mid b$ by definition. If $p \mid a$, we write $a = up$ for some $u \in A$. But then $p = ubp$, i.e., $(1 - ub)p = 0$. Since A is an integral domain and $p \neq 0$, we must have $1 - ub = 0$, i.e., $ub = 1$, i.e., b is a unit. Similarly, $p \mid b$ implies a is a unit. ◀

The converse of this proposition is not true for a general integral domain. Consider the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$. (It is easy to verify that $\mathbb{Z}[\sqrt{-5}]$ is indeed a ring.) $\mathbb{Z}[\sqrt{-5}]$, being a subring of $\mathbb{C}$, is clearly an integral domain. We have two essentially different factorizations of $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. Thus $2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5})$, but $2 \not| (1 + \sqrt{-5})$ and $2 \not| (1 - \sqrt{-5})$, i.e., 2 is not a prime. In order to show that 2 is irreducible let us write $2 = (a + b\sqrt{-5})(c + d\sqrt{-5})$ with $a, b, c, d \in \mathbb{Z}$, so that $4 = (a^2 + 5b^2)(c^2 + 5d^2)$. It is easy to see that the Diophantine equation $x^2 + 5y^2 = 2$ does not have a solution in integer values of x and y. Therefore, we must have $a^2 + 5b^2 = 1$ or $c^2 + 5d^2 = 1$, i.e., $a = \pm 1, b = 0$ or $c = \pm 1, d = 0$, i.e., either $a + b\sqrt{-5}$ or $c + d\sqrt{-5}$ is a unit in $\mathbb{Z}[\sqrt{-5}]$. It can be similarly proved that each of 3, $1 + \sqrt{-5}$ and $1 + \sqrt{-5}$ is also irreducible, but neither is a prime.

Thus $\mathbb{Z}[\sqrt{-5}]$ is an example of a ring where unique factorization fails in some sense. On the other hand, we know that each non-zero integer can be written uniquely as a product of prime integers (the fundamental theorem of arithmetic). In order to make these concepts rigid, we introduce the following definition.

1.5 Definition An integral domain A is called a unique factorization domain or a UFD for short, if every non-zero element $a \in A$ can be written as a product $a = up_1 \cdots p_r$ of primes $p_1, \ldots, p_r$ and a unit u, where the primes p_i (not necessarily all distinct or all non-associate) are uniquely determined by a upto multiplications by units and upto permutations of the indexes $1, \ldots, r$.

By the fundamental theorem of arithmetic $\mathbb{Z}$ is a UFD. So also is the ring $K[X]$ of polynomials over a field K in one indeterminate X. We will later prove both these statements in a rather indirect way. The readers may look at more direct proofs elsewhere. The ring $\mathbb{Z}[\sqrt{-5}]$, on the other hand, is not a UFD, since the element 6 is itself non-zero, non-unit and non-prime and does not admit a factorization into primes. In fact, $\mathbb{Z}[\sqrt{-5}]$ contains irreducible elements which are not primes. In any UFD this is not possible.

1.6 Proposition Let A be a UFD and $p \in A$. Then p is prime, if and only if p is irreducible.

Proof Since a UFD is by definition an integral domain, every prime in A is irreducible by Proposition 1.4. For the converse let p be irreducible and let $p \mid ab$ with $a, b \in A$. We have to show that $p \mid a$ or $p \mid b$. We have $ab = pc$ for some $c \in A$. Consider the unique factorization of $c = up_1 \cdots p_r$ with u a unit and p_i primes in A. Let us write $a_0 := u^{-1}a$ and $b_0 := b$. Then $a_0 b_0 = pp_1 \cdots p_r$. Now p_1 is a prime dividing $a_0 b_0$. Hence we must have $p_1 \mid a_0$ or $p_1 \mid b_0$. If $p_1 \mid a_0$, let us write $a_0 = p_1 a_1$ and $b_1 = b_0$, whereas if

$p_1 \mid b_0$, let us write $a_1 = a_0$ and $b_0 = p_1 b_1$. In both cases canceling (This is allowed, since A is an integral domain.) p_1 gives $a_1 b_1 = pp_2 \cdots p_r$. Continuing this procedure for $p_2, \ldots, p_r$ gives $a_r b_r = p$. But p is irreducible by hypothesis. Therefore, either a_r or b_r is a unit. If a_r is a unit $b_r = a_r^{-1} p$, i.e., $p \mid b_r$. By construction $b_r \mid b_{r-1}$, $b_{r-1} \mid b_{r-2}$, $\ldots$, $b_1 \mid b_0$ and $b_0 \mid b$, i.e., $p \mid b$. Similarly, if b_r is a unit, then $p \mid a$. ◀

Similar (i.e., associate) primes can be grouped together in the factorization of a non-zero element a of a UFD A and we can write $a = up_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where u is a unit, $p_1, \ldots, p_r$ are pairwise non-associate primes and α_i are positive (or nonnegative) integers. Unique factorization means that the primes p_i and the integers α_i are uniquely determined by a (where two associate primes are identified in the notion of uniqueness). This leads us to the following important definitions.

1.7 Definition Let A be a UFD and let $a, b \in A \setminus \{0\}$ admit factorizations of the form $a = up_1^{\alpha_1} \cdots p_r^{\alpha_r}$ and $a = vp_1^{\beta_1} \cdots p_r^{\beta_r}$, where u and v are units, $p_1, \ldots, p_r$ are pairwise non-associate primes and α_i, β_i are nonnegative integers. A greatest common divisor or a gcd of a and b is defined as

$$\gcd(a, b) := w \prod_{i=1}^{r} p_i^{\min(\alpha_i, \beta_i)},$$

where w is any unit of A. Thus $\gcd(a, b)$ is an element of A which is unique upto multiplication by units. In a similar vein a least common multiple or an lcm of a and b is defined as

$$\operatorname{lcm}(a, b) := w' \prod_{i=1}^{r} p_i^{\max(\alpha_i, \beta_i)}$$

for any unit w' of A. Thus $\operatorname{lcm}(a, b)$ is again uniquely determined by a and b upto multiplication by units.

It is often convenient to define $\gcd(a, 0)$ to be an associate of a for every non-zero $a \in A$. However, $\gcd(0, 0)$ and $\operatorname{lcm}(a, 0)$ are left undefined.

If $A = \mathbb{Z}$, the primes of A occurring in the factorizations of nonzero $a \in \mathbb{Z}$ are conventionally taken to be the positive primes. Similarly the positive gcd (resp. lcm) of a and b is usually called *the* gcd (resp. lcm) of a and b. In $K[X]$ too one often makes the gcd and lcm unique by considering the *monic* gcd or lcm of two non-zero polynomials.

1.8 Proposition Let A be a UFD, $0 \neq a, b \in A$, d a gcd of a and b and m an lcm of a and b. If $d' \mid a$ and $d' \mid b$, then $d' \mid d$. Similarly, if $a \mid m'$ and $b \mid m'$, then $m \mid m'$.

Proof This is easy to verify. One may start by showing that if a has a factorization as in Definition 1.7, then all the divisors of A are $u' p_1^{\gamma_1} \cdots p_r^{\gamma_r}$, where u' is a unit and γ_i are integers satisfying $0 \leqslant \gamma_i \leqslant \alpha_i$. ◀

The gcd and lcm of two elements of A have alternate definitions for a special class of UFDs.

1.9 Definition An integral domain A is called a principal ideal domain or a PID in short, if every ideal of A is principal (i.e., generated by a single element).

Again the concepts of prime and irreducible elements in a PID are equivalent.

1.10 Proposition Let A be a PID and $p \in A$. Then p is prime, if and only if p is irreducible.

Proof The 'only if' part is immediate from Proposition 1.4. For the converse let $p \in A$ be irreducible, but not prime. Then there are $a, b \in A$ such that $a \notin \langle p \rangle$ and $b \notin \langle p \rangle$, but $ab \in \langle p \rangle$. Consider the ideal $\mathfrak{a} = \langle \alpha \rangle = \langle p \rangle + \langle a \rangle$. Since $p \in \langle \alpha \rangle$, we have $p = c\alpha$ for some $c \in R$. By hypothesis p is irreducible, so that either c or α is a unit. If c is a unit, $\langle p \rangle = \langle \alpha \rangle = \langle p \rangle + \langle a \rangle$, that is, $a \in \langle p \rangle$, a contradiction. So α is a unit. Then $\langle p \rangle + \langle a \rangle = A$ which implies that there are elements $u, v \in A$ such that $up + va = 1$. Similarly there are elements $u', v' \in A$ such that $u'p + v'b = 1$. Multiplying these two equations gives $(uu'p + uv'b + u'va)p + (vv')ab = 1$. Now $ab \in \langle p \rangle$, so that $ab = wp$ for some $w \in A$. But then $(uu'p + uv'b + u'va + vv'w)p = 1$, which shows that p is a unit, a contradiction. ◀

As claimed earlier, PIDs are UFDs. Proving this requires the following lemma. In Chapter 2 we will generalize and extend it considerably for Noetherian rings. For the time being, the following result is all we need.

1.11 Lemma Let A be a PID. Then we can not have a strictly ascending infinite chain of ideals in A, i.e., we can not have ideals $\mathfrak{a}_1, \mathfrak{a}_2, \ldots$ of A with $\mathfrak{a}_1 \subsetneqq \mathfrak{a}_2 \subsetneqq \cdots$.

Proof Assume that such a chain exists. $\mathfrak{a} := \bigcup_{i=1}^{\infty} \mathfrak{a}_i$ is clearly an ideal of A and is principal by hypothesis. Let $\mathfrak{a} = \langle a \rangle$. Thus $a \in \mathfrak{a}$, i.e., $a \in \mathfrak{a}_i$ for some i. But then for all $j \geqslant i$ we have $\mathfrak{a} \subseteq \mathfrak{a}_j \subseteq \mathfrak{a}$, i.e., $\mathfrak{a}_j = \mathfrak{a}$, a contradiction. ◀

1.12 Theorem A PID is a UFD.

Proof Let A be a PID and $0 \neq a \in A$. We first show that a can be written as a product of finitely many primes. If a is a unit or a prime, we are done. Otherwise, a is not irreducible (Proposition 1.6) and we write $a = a_1 b_1$ with both a_1 and b_1 non-units. If a_1 is prime, it is a prime divisor of a. Otherwise, a_1 is also reducible and we can write $a_1 = a_2 b_2$ with both a_2 and b_2 non-units. If a_2 is prime, it is a prime divisor of a, else we write $a_2 = a_3 b_3$ with both a_3 and b_3 non-units. This process must stop after finitely many steps, i.e., some a_i must be prime, since otherwise we will get an infinite strictly ascending chain $\langle a \rangle \subsetneqq \langle a_1 \rangle \subsetneqq \langle a_2 \rangle \subsetneqq \cdots$. Thus a has one prime divisor; call it p_1 and write $a = p_1 c_1$. If c_1 is a unit, we have produced a prime factorization of a. If c_1 is a non-unit, we find as before a prime divisor p_2 of c_1 and write $c_1 = p_2 c_2$. If c_2 is a unit, we have the factorization $a = p_1 p_2 c_2$, otherwise we write $c_2 = p_3 c_3$ with p_3 prime. Again this process must stop after finitely many steps, i.e., some c_j has to be a unit, since otherwise we will be able to generate an infinite ascending chain $\langle a \rangle \subsetneqq \langle c_1 \rangle \subsetneqq \langle c_2 \rangle \subsetneqq \cdots$.

Let $a = up_1 \cdots p_r = vq_1 \cdots q_s$ be two factorizations of a (with a, b units and p_i, q_j primes). We will show that $r = s$ and $p_i \sim q_{\sigma(i)}$ for some permutation σ of $\{1, 2, \ldots, r\}$. We proceed by induction on $t := \min(r, s)$. If $t = 0$, either $r = 0$ or $s = 0$. If $r = 0$, a is a unit and hence we must have $s = 0$. Similarly, if $s = 0$, we have $r = 0$. In both these cases, the unique factorization of a is obvious. Now assume that the result holds for $t - 1$. Since p_r is a prime, $p_r \mid q_j$ for some j. After rearranging, if necessary, we may take $p_r \mid q_s$, i.e., $q_s = p_r d$. But q_s is prime and hence irreducible, whereas p_r is not a unit. Therefore, d must be a unit, i.e., $p_r \sim q_s$. But then $a' := a/p_r = up_1 \cdots p_{r-1} = (vd)q_1 \cdots q_{s-1}$. By the induction hypothesis, $r - 1 = s - 1$, i.e., $r = s$, and $p_i \sim q_{\sigma(i)}$ for some permutation σ of $\{1, 2, \ldots, r - 1\}$. This completes the proof. ◀

The converse of the last theorem is, however, not true in general. In the exercises we will see that if A is a UFD, so is the polynomial ring $A[X]$ and, in general, the polynomial rings $A[X_1, \ldots, X_n]$. However, if $n \geqslant 2$, then $A[X_1, \ldots, X_n]$ is not a PID, since the ideal $\langle X_1, \ldots, X_n \rangle$ is clearly not principal.

We are now ready to demonstrate the equivalent characterizations of gcd and lcm in a PID.

1.13 Proposition Let A be a PID and $a, b \in A$ not both zero. Let d be a gcd of a and b. Then $\langle d\rangle = \langle a\rangle + \langle b\rangle$. In particular, there exist $u, v \in A$ such that $d = ua + vb$ (Bézout's relation). If a and b are both non-zero and if m is an lcm of a and b, then $\langle m\rangle = \langle a\rangle \cap \langle b\rangle$.

Proof Let $\langle a\rangle + \langle b\rangle = \langle c\rangle$. We show that c and d are associates. There exist $u', v' \in A$ such that $u'a + v'b = c$. Since $d \mid a$ and $d \mid b$, we have $d \mid c$. On the other hand, $a \in \langle c\rangle$, so that $c \mid a$. Similarly $c \mid b$. Proposition 1.8 then implies that $c \mid d$. The characterization of m is similar and left to the reader. ◀

It is, however, not usual (or even efficient) to factorize two integers (or polynomials) for computing their gcd. One uses the so-called Euclidean gcd loop instead. Thus $\mathbb{Z}$ or $K[X]$ are UFDs in which we have something like a Euclidean division algorithm. This is formalized in the following definition.

1.14 Definition An integral domain A is called a Euclidean domain or an ED for brevity, if there exists a map $\nu : A \setminus \{0\} \to \mathbb{Z}_+$ satisfying the following two conditions:

(1) $\nu(a) \leqslant \nu(ab)$ for all $a, b \in A \setminus \{0\}$.

(2) For every $a, b \in A$ with $b \neq 0$ there exist (not necessarily unique) $q, r \in A$ such that $a = qb + r$ with $r = 0$ or $\nu(r) < \nu(b)$.

In this case the map ν is often called a Euclidean degree function. We call q and r respectively a quotient and a remainder of Euclidean division of a by b and denote this as $q = a \operatorname{quot} b$ and $r = a \operatorname{rem} b$.

1.15 Example $\mathbb{Z}$ is an ED with $\nu(a) = |a|$ for $a \neq 0$. For $b \neq 0$ the remainder $r := a \operatorname{rem} b$ (and hence the quotient also) can be made unique by choosing r to be non-negative. The polynomial ring $K[X]$ over a field K is an ED with $\nu(a) = \deg a$ for $a \neq 0$. In this case the remainder $a \operatorname{rem} b$ and the quotient $a \operatorname{quot} b$ are always uniquely determined by a and b.

1.16 Theorem An ED is a PID.

Proof Let A be an ED with Euclidean degree function ν. Obviously the zero ideal of A is principal. Let $\mathfrak{a}$ be a non-zero ideal of A. Choose a to be a non-zero element of $\mathfrak{a}$ such that $\nu(b) \geqslant \nu(a)$ for every $b \in \mathfrak{a}\setminus\{0\}$. We will show that $\mathfrak{a} = \langle a\rangle$. Clearly $\langle a\rangle \subseteq \mathfrak{a}$. For the converse take any $b \in \mathfrak{a}$. By definition there exist $q, r \in A$ with $b = qa + r$ and $r = 0$ or $\nu(r) < \nu(a)$. Then $r = b - qa \in \mathfrak{a}$ and the minimality of $\nu(a)$ forces $r = 0$, i.e., $b = qa \in \langle a\rangle$. ◀

Theorems 1.12 and 1.16 in conjunction with Example 1.15 show that $\mathbb{Z}$ and $K[X]$ (where K is a field) are both PIDs and hence both UFDs. But note that every PID is not necessarily an ED. For example, the ring $\left\{a + b\left(\frac{1+\sqrt{-19}}{2}\right) \mid a, b \in \mathbb{Z}\right\}$ is a PID but not an ED.

The following theorem is at the heart of the Euclidean gcd algorithm.

1.17 Theorem Let A be an ED, $0 \neq a, b \in A$ and $r := a \operatorname{rem} b$. Then $\gcd(a, b) \sim \gcd(b, r)$.

Proof We will show that $\langle a\rangle + \langle b\rangle = \langle b\rangle + \langle r\rangle$. Let $a = qb + r$. Then $a \in \langle b\rangle + \langle r\rangle$ and, therefore, $\langle a\rangle + \langle b\rangle \subseteq \langle b\rangle + \langle r\rangle$. Conversely, $r = a - qb \in \langle a\rangle + \langle b\rangle$ and it follows that $\langle b\rangle + \langle r\rangle \subseteq \langle a\rangle + \langle b\rangle$. ◀

The reader is requested to fill out the details on how this theorem can be used to prove the correctness of the traditional Euclidean gcd algorithm in an ED.

The central object of study in this course is the class of rings called number rings. Number rings are not necessarily unique factorization domains (or PIDs or EDs obviously). However, they are all examples of Dedekind domains which are rings where we have unique factorization of *ideals*. We will study these topics in detail later during this course. The simpler cases of UFDs, PIDS and EDs discussed so far should be the first step of generalizing the notion of unique factorization. Moreover, number rings can be UFDs, PIDs and even EDs. Cool! But it is high time now that we concentrate on ideals in a ring. This is what we do next. Unique factorization is important and if preserving it requires us to play at the ideal-level (instead of the element-level), we will do that.

Exercises for Section 1.1

1. Let A be a ring and $a, b, c, x, y \in A$. Show that:

(a) $a \mid b$, if and only if $\langle a \rangle \supseteq \langle b \rangle$. (In the notation of ideals "divides" means "contains".)

(b) $a \mid b$ and $b \mid a$, if and only if $\langle a \rangle = \langle b \rangle$. Furthermore, if A is an integral domain, then $a \mid b$ and $b \mid a$, if and only if $a \sim b$.

(c) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(d) If $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$.

2. Let A and B be two rings. Show that $(A \times B)^* = A^* \times B^*$.

3. **(a)** Let A be an integral domain. Show that $(A[X])^* = A^*$.

(b) Demonstrate by an example that the result of Part (a) may not hold, if A is *not* an integral domain.

4. Let A be a UFD and $0 \neq a, b, c \in A$. Show that:

(a) $\gcd(a, b) \cdot \operatorname{lcm}(a, b) \sim ab$.

(b) If $a \mid bc$ and $\gcd(a, c)$ is a unit, then $a \mid b$.

5. [Polynomials over a UFD] Let A be a UFD. For a nonzero polynomial $f(X) \in A[X]$ a gcd of the coefficients of $f(X)$ is called a content of $f(X)$ and is denoted by $\operatorname{cont} f(X)$. One can then write $f(X) = (\operatorname{cont} f(X)) f_1(X)$, where $f_1(X) \in A[X]$ with $\operatorname{cont} f_1(X) \in A^*$. $f_1(X)$ is called a primitive part of $f(X)$ and is often denoted as $\operatorname{pp} f(X)$. It is clear that $\operatorname{cont} f(X)$ and $\operatorname{pp} f(X)$ are unique upto multiplication by units of A. If for a non-zero polynomial $f(X) \in A[X]$ the content $\operatorname{cont} f(X) \in A^*$ (or, equivalently, if f and $\operatorname{pp} f(X)$ are associates in $A[X]$), then $f(X)$ is called a primitive polynomial.

Prove Gauss's lemma: For two non-zero polynomials $f(X), g(X) \in A[X]$ the elements $\operatorname{cont}(f(X)g(X))$ and $(\operatorname{cont} f(X))(\operatorname{cont} g(X))$ are associates in A. In particular, the product of two primitive polynomials is again primitive.

6. Let d be a non-zero integer which is not a perfect square. Define $\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}$. Prove the following assertions:

(a) $\mathbb{Z}[\sqrt{d}]$ is an integral domain.

(b) $\mathbb{Z}[\sqrt{d}] \cong \mathbb{Z}[X]/\langle X^2 - d \rangle$.

(c) $(\mathbb{Z}[\sqrt{-1}])^* = \{\pm 1, \pm\sqrt{-1}\}$.

(d) If $d < -1$, then $(\mathbb{Z}[\sqrt{d}])^* = \{\pm 1\}$.

*** (e)** If $d > 0$ and $\mathbb{Z}[\sqrt{d}]$ has a unit other than ± 1, then $\mathbb{Z}[\sqrt{d}]$ has infinitely many units. (**Hint:** Such a unit must be of infinite order.)

*** (f)** For $d = -2, -1, 2, 3$ the ring $\mathbb{Z}[\sqrt{d}]$ is an ED with Euclidean degree function $\nu(a + b\sqrt{d}) = |a^2 - db^2|$, $a, b \in \mathbb{Z}$, not both 0.

*** (g)** If $d < 0$, then $\mathbb{Z}[\sqrt{d}]$ is an ED, if and only if $d = -1$ or -2. (**Hint:** If $d < -2$, you may prove that 2 is irreducible but not prime in $\mathbb{Z}[\sqrt{d}]$. This implies that $\mathbb{Z}[\sqrt{d}]$ is not a UFD, hence not a PID and hence not an ED. Note that $2 \mid d(d-1) = (d + \sqrt{d})(d - \sqrt{d})$.)

7. [Extended Euclidean gcd algorithm] Let A be an ED, $0 \neq a, b \in A$ and let $d := \gcd(a, b)$. Since A is also a PID, there exist $u, v \in A$ such that $d = ua + vb$. Modify the Euclidean gcd algorithm so that it returns all of d, u and v.

1.2 Ideals in a ring

Ideals play a very crucial role in the study of rings (commutative with identity). The concept of ideals was introduced by Ernst Eduard Kummer (1810–1893) and later formalized by Richard Dedekind (1831–1916) in his famous work "Über die Theorie der ganzen algebraischen Zahlen". As before I will use lowercase Gothic letters $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{m}, \mathfrak{n}, \mathfrak{p}, \mathfrak{q}$ (respectively, a, b, c, m, n, p, q) to designate ideals. On some specific occasions I will also use the upper-case Gothic letters $\mathfrak{J}, \mathfrak{N}, \mathfrak{P}$ and $\mathfrak{Q}$ (J, N, P and Q) to denote certain ideals. Mathematicians always run out of symbols and many believe that if it is Gothic, it looks ideal.

We start with some basic operations on ideals.

1.18 Definition Let A be a ring and let $\mathfrak{a}_i$, $i \in I$, be a family (not necessarily finite) of ideals in A.

- The set-theoretic intersection $\bigcap_{i\in I} \mathfrak{a}_i$ is evidently an ideal in A.
- The sum of the family $\mathfrak{a}_i$ is the ideal

$$\sum_{i\in I} \mathfrak{a}_i := \left\{ \sum_{i\in I} x_i \mid x_i \in \mathfrak{a}_i \text{ and } x_i = 0 \text{ except for finitely many } i \in I \right\}.$$

  Two ideals $\mathfrak{a}$ and $\mathfrak{b}$ of A are said to be relatively prime or coprime, if $\mathfrak{a} + \mathfrak{b} = A$, or equivalently if there exist $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ with $a + b = 1$.

- If $I = \{1, 2, \ldots, n\}$ is *finite*, the product $\mathfrak{a}_1\mathfrak{a}_2 \cdots \mathfrak{a}_n$ is the ideal generated by all elements of the form $x_1x_2 \cdots x_n$ with $x_i \in \mathfrak{a}_i$ for all $i = 1, \ldots, n$. It is easy to see that

$$\mathfrak{a}_1\mathfrak{a}_2 \cdots \mathfrak{a}_n = \left\{ \sum_{j=1}^{r} x_{j,1}x_{j,2} \cdots x_{j,n} \mid r \in \mathbb{Z}_+, x_{j,i} \in \mathfrak{a}_i \right\}.$$

  If $\mathfrak{a}_1 = \mathfrak{a}_2 = \cdots = \mathfrak{a}_n = \mathfrak{a}$, the product $\mathfrak{a}_1\mathfrak{a}_2 \cdots \mathfrak{a}_n$ is often denoted as $\mathfrak{a}^n$.

It is easy to see that the operations intersection, sum and product on ideals in a ring are associative and commutative.

1.19 Theorem [Chinese remainder theorem (CRT)] Let A be a ring and $n \in \mathbb{N}$. Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ be ideals in A such that for all i, j, $i \neq j$, the ideals $\mathfrak{a}_i$ and $\mathfrak{a}_j$ are relatively prime. Then $A/(\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n)$ is isomorphic to the direct product $A/\mathfrak{a}_1 \times \cdots \times A/\mathfrak{a}_n$.

Proof The assertion is obvious for $n = 1$. So assume that $n \geqslant 2$ and define the map $\varphi : A/(\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n) \to A/\mathfrak{a}_1 \times \cdots \times A/\mathfrak{a}_n$ by $a + (\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n) \mapsto (a + \mathfrak{a}_1, \ldots, a + \mathfrak{a}_n)$ for all $a \in A$. Since $\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n \subseteq \mathfrak{a}_i$ for all i, the map is well-defined. It is easy to see that φ is a ring homomorphism. In order to show that φ is injective, we let $\varphi(a + (\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n)) = 0$. This means that $a + \mathfrak{a}_i = 0$, that is, $a \in \mathfrak{a}_i$ for all i. Then $a \in \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n$, that is, $a + (\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n) = 0$.

The trickier part is to prove that φ is surjective. Let $a_1, \ldots, a_n \in A$. Let us consider the ideals $\mathfrak{b}_i := \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_{i-1} \cap \mathfrak{a}_{i+1} \cap \cdots \cap \mathfrak{a}_n$ for each i. For a given i there exist for each $j \neq i$ elements $\alpha_j \in \mathfrak{a}_i$ and $\beta_j \in \mathfrak{a}_j$ such that $\alpha_j + \beta_j = 1$. Multiplying these equations shows that we have a $\gamma_i \in \mathfrak{a}_i$ such that $\gamma_i + \delta_i = 1$, where $\delta_i := \beta_1 \cdots \beta_{i-1}\beta_{i+1} \cdots \beta_n \in \mathfrak{b}_i$. (This shows that $\mathfrak{a}_i + \mathfrak{b}_i = A$ for all i.) Now consider the element $a := \sum_{i=1}^{n} \delta_i a_i$. It follows that $a \equiv a_i \pmod{\mathfrak{a}_i}$ for all i, that is, $\varphi(a + (\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n)) = (a_1 + \mathfrak{a}_1, \ldots, a_n + \mathfrak{a}_n)$. ◀

Note that in the proof of the last theorem φ is injective unconditionally, i.e., for any ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$. The pairwise coprimality of these ideals has been needed only to prove the surjectivity of φ.

1.20 Corollary Let $m_1, \ldots, m_n \in \mathbb{N}$ be pairwise relatively prime moduli. Then for integers $a_1, \ldots, a_n$ there exists an integer a unique modulo $m_1 \cdots m_n$ such that $a \equiv a_i \pmod{m_i}$ for all $i = 1, \ldots, n$. ◀

Two particular types of ideals are very important in algebra.

1.21 Definition Let A be a ring.

- An ideal $\mathfrak{p}$ of A is called a `prime ideal`, if $\mathfrak{p} \neq A$ and if $ab \in \mathfrak{p}$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$ for $a, b \in A$. The second condition is equivalent to saying that if $a \notin \mathfrak{p}$ and $b \notin \mathfrak{p}$, then the product $ab \notin \mathfrak{p}$.
- An ideal $\mathfrak{m}$ of A is called a `maximal ideal`, if $\mathfrak{m} \neq A$ and if for any ideal $\mathfrak{a}$ satisfying $\mathfrak{m} \subseteq \mathfrak{a} \subseteq A$ we have $\mathfrak{a} = \mathfrak{m}$ or $\mathfrak{a} = A$. The second condition means that there are no non-unit ideals of A properly containing $\mathfrak{m}$.

1.22 Example For $p \in \mathbb{P}$ the principal ideal $\langle p \rangle$ of $\mathbb{Z}$ is prime. On the other hand, for a composite integer n the ideal $\langle n \rangle$ of $\mathbb{Z}$ is not prime. For example, $2 \notin \langle 6 \rangle$ and $3 \notin \langle 6 \rangle$, but the product $2 \times 3 \in \langle 6 \rangle$.

The ideal $\langle p \rangle$ of $\mathbb{Z}$ for a prime p is also maximal, for if $\langle p \rangle \subsetneqq \mathfrak{a} \subseteq \mathbb{Z}$ ($\mathfrak{a}$ an ideal in $\mathbb{Z}$), then $\mathfrak{a}$ contains an integer a which is not a multiple of and hence is coprime to p. By Bézout's theorem there exist integers u, v with $up + va = 1$ implying that $1 \in \mathfrak{a}$, i.e., $\mathfrak{a} = \mathbb{Z}$.

Next consider the polynomial ring $A = \mathbb{Z}[X]$ and the principal ideal $\langle X \rangle$ of A. It is easy to see that $\langle X \rangle \subsetneqq \langle X, 2 \rangle \subsetneqq A$. Thus $\langle X \rangle$ is not maximal.

Prime and maximal ideals can be characterized by the following equivalent criteria.

1.23 Proposition Let A be a ring and $\mathfrak{a}$ an ideal of A.

(1) $\mathfrak{a}$ is a prime ideal of A, if and only if $A/\mathfrak{a}$ is an integral domain. In particular, A is an integral domain, if and only if the zero ideal of A is prime.

(2) $\mathfrak{a}$ is a maximal ideal of A, if and only if $A/\mathfrak{a}$ is a field.

Proof (1) Let $a, b \in A$ be arbitrary. Then $\mathfrak{a}$ is prime $\iff$ $ab \in \mathfrak{a}$ implies $a \in \mathfrak{a}$ or $b \in \mathfrak{a}$ $\iff$ $ab + \mathfrak{a} = (a + \mathfrak{a})(b + \mathfrak{a}) = 0$ implies $a + \mathfrak{a} = 0$ or $b + \mathfrak{a} = 0$ $\iff$ $A/\mathfrak{a}$ is an integral domain.

(2) Let $\mathfrak{a}$ be a maximal ideal. Choose $b + \mathfrak{a} \neq 0 + \mathfrak{a}$. Then $b \notin \mathfrak{a}$. Consider the ideal $\mathfrak{b} := \mathfrak{a} + \langle b \rangle$. Since $\mathfrak{a}$ is maximal, we must have $\mathfrak{b} = A$. This means that $a + cb = 1$ for some $a \in \mathfrak{a}$ and $c \in A$. Then $(c + \mathfrak{a})(b + \mathfrak{a}) = 1 + \mathfrak{a}$ which implies that $b + \mathfrak{a}$ is a unit in $A/\mathfrak{a}$. That is, $A/\mathfrak{a}$ is a field.

Conversely, let $A/\mathfrak{a}$ be a field. Consider any ideal $\mathfrak{b}$ of A with $\mathfrak{a} \subsetneqq \mathfrak{b} \subseteq A$. Choose any $b \in \mathfrak{b} \setminus \mathfrak{a}$. Then $b + \mathfrak{a} \neq 0 + \mathfrak{a}$. By hypothesis there exists $c \in R$ such that $(b + \mathfrak{a})(c + \mathfrak{a}) = 1 + \mathfrak{a}$, that is, $bc - 1 \in \mathfrak{a} \subseteq \mathfrak{b}$. Hence $1 \in \mathfrak{b}$, that is, $\mathfrak{b} = A$. ◀

Since fields are integral domains, we immediately have the following important corollary.

1.24 Corollary Maximal ideals are prime. ◀

A question that naturally arises is whether there exist prime and maximal ideals in every ring. The answer is affirmative as long as the ring is non-zero. The following proof is an interesting application of Zorn's lemma[1].

1.25 Proposition Every non-zero ring A has at least one maximal (and hence at least one prime) ideal.

Proof Let S denote the set of all proper ideals of A. S is non-empty, since the zero ideal of A is a proper ideal of $A \neq 0$. S is clearly a partially ordered set under the relation $\subseteq$ (inclusion). Now let $T = \{\mathfrak{a}_i \mid i \in I\}$ be a chain in S. Consider $\mathfrak{a} := \bigcup_{i\in I} \mathfrak{a}_i$ and take $a, b \in \mathfrak{a}$. Then $a \in \mathfrak{a}_i$ and $b \in \mathfrak{a}_j$ for some $i, j \in I$. Since either $\mathfrak{a}_i \subseteq \mathfrak{a}_j$ or $\mathfrak{a}_j \subseteq \mathfrak{a}_i$, it follows that both a and b belong to either $\mathfrak{a}_i$ or $\mathfrak{a}_j$ and, therefore, $a \pm b$, ca (for any $c \in A$) are in either $\mathfrak{a}_i$ or $\mathfrak{a}_j$ and hence in $\mathfrak{a}$. Thus $\mathfrak{a}$ is an ideal of A. Furthermore, $1 \notin \mathfrak{a}_i$ for all $i \in I$ and hence $1 \notin \mathfrak{a}$, i.e., $\mathfrak{a} \in S$. Thus the chain T has an upper bound in S. By Zorn's lemma, S has at least one maximal element. ◀

The last proof can be easily modified to derive the following more general result. Alternatively, one may use Proposition 1.25 on $A/\mathfrak{a}$ and the one-to-one correspondence between the ideals in $A/\mathfrak{a}$ and those of A containing $\mathfrak{a}$.

1.26 Proposition Let A be a ring and $\mathfrak{a}$ a *proper* ideal of A. Then there exists at least one maximal (and hence at least one prime) ideal of A containing $\mathfrak{a}$. In particular, for every non-unit a of A there exists a maximal (and hence a prime) ideal of A containing a. ◀

The set of all prime ideals in A is called the (prime) spectrum of A and is denoted by $\operatorname{Spec} A$. Similarly, the set of all maximal ideals of A is called the maximal spectrum of A and denoted by $\operatorname{Spm} A$. We have $\operatorname{Spm} A \subseteq \operatorname{Spec} A$. Furthermore, if A is non-zero, both these sets are non-empty. In modern algebra these two sets play an extremely useful role for the study of the ring A. For example, one can define a topology (the so-called Zariski topology) on $\operatorname{Spec} A$ and $\operatorname{Spm} A$ naturally inherits its share from $\operatorname{Spec} A$. Grothendieck's language of schemes exploits the structures of these sets and provides a unifying ground for algebraic geometry and algebraic number theory. But these modern languages of mathematics are little too abstract and advanced to be included in a simple-minded course like this. Interested students may study these topics later in their careers.

In what follows I will use the Gothic letters $\mathfrak{p}, \mathfrak{q}, \mathfrak{P}, \mathfrak{Q}$ to denote prime ideals, and the letter $\mathfrak{m}$ to denote maximal ideals.

1.27 Definition Let A be a non-zero ring. The nilradical of A is defined to be the ideal

$$\mathfrak{N}_A := \bigcap_{\mathfrak{p} \in \operatorname{Spec} A} \mathfrak{p},$$

i.e., to be the intersection of all prime ideals of A. Similarly, the Jacobson radical $\mathfrak{J}_A$ is the intersection of all maximal ideals of A, or in other words, it is the ideal defined by

$$\mathfrak{J}_A := \bigcap_{\mathfrak{m} \in \operatorname{Spm} A} \mathfrak{m}.$$

[1]Let S be a non-empty set partially ordered by the relation $\leqslant$. A chain of S is a subset T of S such that for any $a, b \in T$ we have either $a \leqslant b$ or $b \leqslant a$. An upper bound of a subset T of S is an element $b \in S$ satisfying $a \leqslant b$ for all $a \in T$. A maximal element of S is an element $c \in S$ such that $c \leqslant a$ for some $a \in S$ implies that $a = c$. Zorn's lemma states that if *every* chain of S has an upper bound in S, then S contains at least one maximal element. Zorn's lemma can not be proved independently, but can be shown to be equivalent to the other axioms of mathematics, like the well-ordering principle or the axiom of choice.

Clearly, $\mathfrak{N}_A \subseteq \mathfrak{J}_A$. When no confusions are likely, we may drop the prefix A from $\mathfrak{N}_A$ and $\mathfrak{J}_A$. If A is the zero ring, it is customary to define $\mathfrak{N}_A = \mathfrak{J}_A = 0$.

Jacobson radical is named after Nathan Jacobson (1910–1999) who contributed greatly to the study of rings, Lie algebras and Jordan algebras. The name 'nilradical' comes from the following consideration. Recall the an element a in a ring A is called nilpotent, if $a^r = 0$ for some $r \in \mathbb{N}$. Clearly, 0 is a nilpotent element in any ring. An example of a non-zero nilpotent element is $[2]_4$ in the ring $\mathbb{Z}_4$. The following result gives a connection between the nilpotent elements and the nilradical. It is left to the reader as an exercise to prove that the set of all nilpotent elements in a ring A is an ideal in A. (Also look at Exercise 1.2.9.)

1.28 Proposition Let A be a ring. Then $\mathfrak{N}_A$ is the set of all nilpotent elements of A.

Proof The result is obvious for $A = 0$. So assume that A is a non-zero ring and let $\mathfrak{n}_A$ denote the set of all nilpotent elements of A. We will show that $\mathfrak{n}_A = \mathfrak{N}_A$. First note that if $a \in \mathfrak{n}_A$, then $a^r = 0$ for some $r \in \mathbb{N}$, i.e., $a^r \in \mathfrak{p}$ for every $\mathfrak{p} \in \operatorname{Spec} A$. Since all such $\mathfrak{p}$ are prime, it follows that $a \in \mathfrak{p}$ for all $\mathfrak{p} \in \operatorname{Spec} A$, i.e., $a \in \mathfrak{N}_A$, so that $\mathfrak{n}_A \subseteq \mathfrak{N}_A$.

For proving the reverse inclusion take any $a \notin \mathfrak{n}_A$. Let S be the set of all proper ideals of A not containing any power a^n, $n \in \mathbb{N}$, of a. Since a is not nilpotent, the zero ideal belongs to S, i.e., S is nonempty. Also S is partially ordered under inclusion $\subseteq$. As in the proof of Proposition 1.25 one can show that S has a maximal element, say, $\mathfrak{p}$. If we can show that $\mathfrak{p} \in \operatorname{Spec} A$, we will have $a \notin \mathfrak{N}_A$, since by construction $\mathfrak{p}$ does not contain any power of a and, in particular, a itself. So take $b, c \notin \mathfrak{p}$. Then the ideals $\mathfrak{p} + \langle b \rangle$ and $\mathfrak{p} + \langle c \rangle$ are strict supersets of $\mathfrak{p}$ and hence by the maximality of $\mathfrak{p}$ are not in S, i.e., contain some powers of a. Let $a^m \in \mathfrak{p} + \langle b \rangle$ and $a^n \in \mathfrak{p} + \langle c \rangle$, i.e., $a^m = u + u'b$ and $a^n = v + v'c$ for some $u, v \in \mathfrak{p}$ and $u', v' \in A$. But then $a^{m+n} = (uv + uv'c + u'vb) + (u'v')bc \in \mathfrak{p} + \langle bc \rangle$, i.e., $\mathfrak{p} + \langle bc \rangle \notin S$, i.e., $bc \notin \mathfrak{p}$. Thus $\mathfrak{p}$ is prime, as desired. ◀

A similar characterization of the Jacobson radical $\mathfrak{J}_A$ is covered in Exercise 1.2.11.

Exercises for Section 1.2

1. **(a)** Let $\mathfrak{a}$ be an ideal in a ring A. Show that $\mathfrak{a} = A$, if and only if $\mathfrak{a}$ contains a unit. In particular, if $\mathfrak{a} \subsetneq A$ (i.e., if $\mathfrak{a}$ is a proper ideal of A), then $\mathfrak{a}$ consists only of non-units. (**Remark:** This is why $A = \langle 1 \rangle$ is called the unit ideal.)

(b) Show that the only ideals in a field are the zero ideal and the unit ideal.

2. Let A be a ring and $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ ideals in A. Prove that:

(a) $\mathfrak{a}(\mathfrak{b} + \mathfrak{c}) = \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{c}$.

(b) $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b}$. Furthermore, if $\mathfrak{a} + \mathfrak{b} = A$, then $\mathfrak{a}\mathfrak{b} = \mathfrak{a} \cap \mathfrak{b}$.

(c) More generally, show that if $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ are pairwise relatively prime ideals of A, then $\mathfrak{a}_1 \cdots \mathfrak{a}_r = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_r$. (**Hint:** Induction on r.)

3. Let A be a ring and $a \in A$. Show that a is prime, if and only if the principal ideal $\langle a \rangle$ of A is prime.

4. Let $f : F \to K$ be a homomorphism of fields. Show that f is injective.

5. Show that a finite integral domain is a field.

6. Let $f : A \to B$ be a ring homomorphism.

(a) Let $\mathfrak{b}$ be an ideal in B and $\mathfrak{a} := f^{-1}(\mathfrak{b}) := \{a \in A \mid f(a) \in \mathfrak{b}\}$. Show that $\mathfrak{a}$ is an ideal in A. Show further that if $\mathfrak{b}$ is prime, then $\mathfrak{a}$ is also prime, but if $\mathfrak{b}$ is maximal, then $\mathfrak{a}$ need not be maximal. (**Remark:** The ideal $\mathfrak{a}$ is called the contraction of $\mathfrak{b}$ and is denoted by $\mathfrak{a} = \mathfrak{b}^c$.)

(b) Let $\mathfrak{a}$ be an ideal of A. Show that $f(\mathfrak{a}) := \{f(a) \mid a \in \mathfrak{a}\}$ is not necessarily an ideal in B. (**Remark:** The ideal generated by the elements of $f(\mathfrak{a})$ is called the extension of $\mathfrak{a}$ and is denoted by $\mathfrak{a}^e$.)

(c) Let $\mathfrak{a}$ be an ideal of A and $\mathfrak{b}$ an ideal of B. Show that: $\mathfrak{a} \subseteq \mathfrak{a}^{ec}$, $\mathfrak{b} \supseteq \mathfrak{b}^{ce}$, $\mathfrak{a}^e = \mathfrak{a}^{ece}$ and $\mathfrak{b}^c = \mathfrak{b}^{cec}$.

7. **(a)** Show that the set $\mathfrak{N}$ of nilpotent elements in a ring A is an ideal of A.

(b) Prove that the ring $A/\mathfrak{N}$ does not contain non-zero nilpotent elements. (**Remark:** One often calls the ring $A/\mathfrak{N}$ the reduction of A and denote this ring as A_{red}. If $\mathfrak{N} = 0$, then $A_{\text{red}} = A$ and we call A to be a reduced ring. In particular, the reduction A_{red} of any ring A is reduced.)

(c) Conclude that an integral domain is reduced. Give an example of a reduced ring that is *not* an integral domain.

8. **(a)** Show that all non-zero prime ideals in a PID are maximal. (**Hint:** Example 1.22.)

(b) Demonstrate by an example that a prime ideal in a UFD need not be maximal.

9. Let A be a ring and $\mathfrak{a} \subseteq A$ an ideal. Consider the set

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

Show that $\sqrt{\mathfrak{a}}$ is an ideal of A. It is called the radical or root of $\mathfrak{a}$. If $\sqrt{\mathfrak{a}} = \mathfrak{a}$, then $\mathfrak{a}$ is called a radical ideal or a root ideal. For arbitrary ideals $\mathfrak{a}$ and $\mathfrak{b}$ of A prove the following assertions.

(a) $\mathfrak{a} \subseteq \sqrt{\mathfrak{a}}$.

(b) $\sqrt{\sqrt{\mathfrak{a}}} = \sqrt{\mathfrak{a}}$.

(c) If $\mathfrak{a} \subseteq \mathfrak{b}$, then $\sqrt{\mathfrak{a}} \subseteq \sqrt{\mathfrak{b}}$.

(d) If $\mathfrak{a}$ is a prime ideal, then $\sqrt{\mathfrak{a}} = \mathfrak{a}$.

(e) $\sqrt{\mathfrak{a}} = A$, if and only if $\mathfrak{a} = A$.

(f) $\sqrt{\mathfrak{a} + \mathfrak{b}} = \sqrt{\sqrt{\mathfrak{a}} + \sqrt{\mathfrak{b}}}$.

(g) $\sqrt{\mathfrak{a} \cap \mathfrak{b}} = \sqrt{\mathfrak{a}} \cap \sqrt{\mathfrak{b}}$.

(h) The nilradical $\mathfrak{N}_A = \sqrt{0}$.

10. [The prime avoidance lemma] Let A be a ring, $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ prime ideals in A and $\mathfrak{a}$ an ideal of A with $\mathfrak{a} \subseteq \bigcup_{i=1}^{n} \mathfrak{p}_i$. Show that $\mathfrak{a} \subseteq \mathfrak{p}_i$ for some $i \in \{1, \ldots, n\}$. (**Hint:** Prove the contrapositive by induction on n.)

11. Let A be a ring and $\mathfrak{J}$ the Jacobson radical of A. Show that $a \in \mathfrak{J}$, if and only if $1 - ab \in A^*$ for all $b \in A$.

1.3 Modules and algebras

Vector spaces and linear transformations between them are the central objects of study in linear algebra. We now generalize the concept of vector spaces to get a more powerful class of objects called modules. A module which also carries a (compatible) ring structure is referred to as an algebra. Algebras over fields (or more generally over rings) play an important role in commutative algebra, algebraic geometry and algebraic number theory.

Recall that a vector space over a field K is an Abelian group V together with a scalar multiplication map $\cdot : K \times V \to V$ enjoying certain properties (loosely speaking, the linearity properties). If we simply assume that K is a general ring (i.e., not necessarily a field) and keep the other parts of the definition (of a vector space) intact, we get a K-module. Now K being a ring, it is not expected in general that every non-zero element in K is invertible. This means that all the properties of vector spaces do not straightaway carry over to modules. But that does not deter us from going for the generalization.

1.29 Definition Let A be a ring. A module over A (or an A-module, in short) is an (additively written) Abelian group M together with a scalar multiplication map $\cdot : A \times M \to M$ with the

following properties. Conventionally we denote $\cdot(a,x)$ as $a\cdot x$ or simply as ax. For every $a,b\in A$ and $x,y\in M$ the scalar multiplication map must satisfy:

$$\begin{aligned} a(x+y) &= ax+ay,\\ (a+b)x &= ax+bx,\\ 1\cdot x &= x,\\ a(bx) &= (ab)x, \end{aligned}$$

where ab denotes the product of a and b in the ring A.

1.30 Example (1) When A is a field, an A-module is a vector space over A.

(2) Ideals of A are modules over A with the ring multiplication taken as the scalar multiplication map.

(3) Every Abelian group G is a $\mathbb{Z}$-module under the scalar multiplication map defined as

$$nx := \begin{cases} 0 & \text{if } n=0\\ x+\cdots+x\ (n \text{ times}) & \text{if } n>0\\ -x-\cdots-x\ (-n \text{ times}) & \text{if } n<0\ . \end{cases}$$

(4) The polynomial rings $A[X]$ and $A[X_1,\ldots,X_n]$ are modules over A.

(5) Let $A\subseteq B$ be any extension of rings. Then B is an A-module with the scalar multiplication map defined as the multiplication of the ring B.

(6) Let M_i, $i\in I$, be a family of A-modules. The direct product $\prod_{i\in I}M_i$ of M_i is defined as the set of all tuples $(x_i)_{i\in I}$ with $x_i\in M_i$. The direct sum $\bigoplus_{i\in I}M_i$ is the subset of the Cartesian product $\prod_{i\in I}M_i$ consisting only of the tuples $(a_i)_{i\in I}$ for which $a_i=0$ except for a finite number of $i\in I$. Both the direct product and the direct sum are A-modules under component-wise addition and scalar multiplication.

If $M_i=M$ for all $i\in I$, we denote the direct product of M_i, $i\in I$, as M^I and the direct sum of M_i, $i\in I$, as $M^{(I)}$. When I is of finite cardinality n, these two modules are naturally the same and we use the notation M^n to designate $M^I=M^{(I)}$ in this case.

The above example shows that all vector spaces. ideals and Abelian groups are modules. This means that any result we prove for modules holds for all these three kinds of algebraic structures. This is one of the reasons why modules call for specific attention in mathematics.

1.31 Proposition Let M be an A-module. Then for every $a\in A$ and $x\in M$ we have: $0\cdot x=0$, $a\cdot 0=0$, $(-a)x=a(-x)=-(ax)$ and $(-a)(-x)=ax$.

Proof Easy verification. ◀

An A-submodule of an A-module M is a subgroup N of M that is closed under the scalar multiplication of M. For an arbitrary subset $S\subseteq M$ the set of all finite linear combinations of the form $a_1x_1+\cdots+a_nx_n$, $n\in\mathbb{Z}_+$, $a_i\in A$, $x_i\in S$, is an A-submodule N of M and is denoted by AS or $\sum_{x\in S}Ax$. We say that N is generated by S (or by elements of S). If S is finite, then N is said to be finitely generated. A (sub)module generated by a single element is called cyclic.

It is important to note that unlike vector spaces the cardinality of a minimal generating set of a module is not necessarily unique. (See Exercise 1.3.1 for an example.) It is also true that given a minimal generating set S for M there may be several different ways of writing a given element of M as finite linear combinations of

elements of S. (For example, if $M = A = \mathbb{Z}$ and $S = \{2, 3\}$, then $1 = (-1) \cdot 2 + 1 \cdot 3 = 2 \cdot 2 + (-1) \cdot 3$.) It follows that the nice theory of dimensions enjoyed by vector spaces does not generalize to modules.

If M is an A-module and N a submodule of M, then the Abelian group M/N can be made to an A-module by defining the scalar multiplication map $a(x + N) := ax + N$. This module (still denoted as M/N) is called the quotient module of M by N. (Note that M is by definition an *Abelian* group, so that any subgroup N of M is normal in M and the quotient group M/N is necessarily defined.)

For A-modules M and N an A-linear map or an A-module homomorphism (from M to N) is defined as a map $f : M \to N$ satisfying $f(ax + by) = af(x) + bf(y)$ for all $a, b \in A$ and $x, y \in M$ (or equivalently satisfying $f(x + y) = f(x) + f(y)$ and $f(ax) = af(x)$ for all $a \in A$ and $x, y \in M$). An isomorphism of modules is a bijective homomorphism. If $M = N$, then an A-linear map $f : M \to M$ is also called an endomorphism of M. Finally, an automorphism is a bijective endomorphism. The set of all (A-module) homomorphisms $M \to N$ is denoted by $\mathrm{Hom}_A(M, N)$ and the set of all (A-module) endomorphisms of M is denoted by $\mathrm{End}_A M$. These sets are again A-modules under the definitions: $(f + g)(x) := f(x) + g(x)$ and $(af)(x) := af(x)$ for all $a \in A$ and $x \in M$ (and f, g in $\mathrm{Hom}_A(M, N)$ or $\mathrm{End}_A M$).

For an A-linear map $f : M \to N$ the kernel and image of f are defined respectively as the sets

$$\mathrm{Ker}\, f := \{x \in M \mid f(x) = 0\} \subseteq M$$

and

$$\mathrm{Im}\, f := \{y \in N \mid y = f(x) \text{ for some } x \in M\} \subseteq N\,.$$

Like groups, rings and vector spaces, we have the isomorphism theorem for modules.

1.32 Theorem [Isomorphism theorem] For an A-module homomorphism $f : M \to N$ the sets $\mathrm{Ker}\, f$ and $\mathrm{Im}\, f$ are submodules of M and N respectively and $M/\mathrm{Ker}\, f \cong \mathrm{Im}\, f$. ◀

Certain specific A-modules behave like vector spaces in the sense that they have bases over A. These modules are worth investigating, because the number rings are $\mathbb{Z}$-modules of this type.

1.33 Definition A free module M over a ring A is defined to be a direct sum $\bigoplus_{i \in I} M_i$ of A-modules M_i with each $M_i \cong A$ as an A-module. Thus a free A-module M is isomorphic to $A^{(I)}$ for some index set I. If I is of finite cardinality n, then $M \cong A^n$.

For example, any vector space over a field K, being isomorphic to $K^{(I)}$ for some index set I, is a free K-module. In particular, every K-vector space of finite dimension n is isomorphic to K^n. On the other hand, every finitely generated A-module need not be isomorphic to a free module A^n for some $n \in \mathbb{N}$. For example, consider an integer $m \geqslant 2$ and the cyclic $\mathbb{Z}$-module $\mathbb{Z}_m$ which has cardinality m. The free $\mathbb{Z}$-modules $\mathbb{Z}^r$ have cardinalities 1 or infinity. Hence $\mathbb{Z}_m$ can not be a free $\mathbb{Z}$-module. We, however, have the following result:

1.34 Theorem [Structure theorem for finitely generated modules] Let M be an A-module. Then M is finitely generated (as an A-module), if and only if M is the quotient of a free module A^n for some $n \in \mathbb{Z}_+$.

Proof [if] The free A-module A^n has a generating set $\{e_1, e_2, \ldots, e_n\}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (1 in the i-th position). If $M = A^n/N$ for some A-submodule N of A^n, the equivalence classes $e_i + N$, $i = 1, \ldots, n$, clearly constitute a finite (but not necessarily minimal) set of generators of M.

[only if] If $x_1, \ldots, x_n$ generate M, then the A-linear map $f : A^n \to M$ defined by $(a_1, \ldots, a_n) \mapsto a_1x_1 + \cdots + a_nx_n$ is surjective. Hence by the isomorphism theorem $M \cong A^n/\operatorname{Ker} f$. ◀

As in the case of vector spaces, a subset S of an A-module M is called linearly independent over A, if an arbitrary finite A-linear combination $\sum_{i=1}^n a_ix_i$ (with $a_i \in A$ and $x_i \in S$) is zero only for $a_1 = \cdots = a_n = 0$. A subset S of M is called an A-basis of M, if S is linearly independent over A and generates M as an A-module. It is easy to see that $S \subseteq M$ is an A-basis of M, if and only if every $x \in M$ can be written *uniquely* as an A-linear combination $x = a_1x_1 + \cdots + a_nx_n$ with $n \in \mathbb{Z}_+$, $a_i \in A$ and $x_i \in S$.

1.35 Proposition Let M be an A-module. Then M has an A-basis, if and only if M is a free A-module.

Proof [if] Let $f : A^{(I)} \to M$ be an isomorphism for some index set I. For each $i \in I$ define $e_i := (\delta_{ij})_{j\in I}$, where δ_{ij} is the Dirac delta. It is easy to see that e_i, $i \in I$, form an A-basis of $A^{(I)}$. It follows that $f(e_i)$, $i \in I$, form an A-basis of M.

[only if] Let x_i, $i \in I$, be an A-basis of M. Define $f : A^{(I)} \to M$ by $(a_i)_{i\in I} \mapsto \sum_{i\in I} a_ix_i$. First note that elements of $A^{(I)}$ are I-tuples $(a_i)_{i\in I}$ with only finitely many a_i non-zero. Therefore, $f((a_i)_{i\in I})$ is well-defined (i.e., a finite sum). It is easy to check that f is an A-linear map. Also f is surjective, since x_i, $i \in I$, is a generating set for M. Furthermore, f is injective, because x_i, $i \in I$, are linearly independent over A. Therefore, f is an isomorphism. ◀

That free modules have bases does not immediately imply that any two bases of a free module will have to have the same cardinality. This is, however, true, though proving this requires some care. We start with the following lemma the (easy) verification of which is left to the reader.

1.36 Lemma For an ideal $\mathfrak{a}$ of A and an A-module M the set $\mathfrak{a}M$ consisting of all finite A-linear combinations $\sum_{i=1}^n a_ix_i$ with $n \in \mathbb{Z}_+$, $a_i \in \mathfrak{a}$ and $x_i \in M$ is an A-submodule of M. $M/\mathfrak{a}M$ is an $A/\mathfrak{a}$-module. Moreover, if M is a free A-module with a basis x_i, $i \in I$, then $M/\mathfrak{a}M$ is also a free $A/\mathfrak{a}$-module with basis $\pi(x_i)$, $i \in I$, where $\pi : M \to M/\mathfrak{a}M$ is the canonical projection map. ◀

Now we can state and prove the dimension theorem for free modules.

1.37 Theorem Let A be a non-zero ring and M a free A-module. Then every A-basis of M has the same cardinality.

Proof Let x_i, $i \in I$, constitute an A-basis of M and let $\mathfrak{a}$ be a *maximal* ideal of A. (Such an ideal exists by Proposition 1.25.) By the last lemma $M/\mathfrak{a}M$ is a free $A/\mathfrak{a}$-module with basis $\pi(x_i)$, $i \in I$, where $\pi : M \to M/\mathfrak{a}M$ is the canonical projection map. But $\mathfrak{a}$ is a maximal ideal of A and hence by Proposition 1.23 $K := A/\mathfrak{a}$ is a field, that is, $M/\mathfrak{a}M$ is a K-vector space. By the dimension theorem for vector spaces we have $|I| = \dim_K(M/\mathfrak{a}M)$. ◀

The cardinality of any basis of a free A-module M is called the rank of M and is denoted by $\operatorname{Rank}_A M$ or simply by $\operatorname{Rank} M$ (if A is understood from the context). This terminology may be a bit confusing, because when vector spaces are concerned we prefer to call dimension instead of rank. On the other hand, we associate the term rank with a linear transformation (or matrix). Note that a vector space can not have that rank which is defined for linear transformations. Thus whenever we say rank of a K-vector space V, we mean the rank of V as a K-module, i.e., $\dim_K V$.

Next we prove a result which turns out to be very useful one.

1.38 Proposition [Nakayama's lemma] Let M be a *finitely generated* A-module and $\mathfrak{a}$ an ideal of A contained in the Jacobson radical $\mathfrak{J}_A$ of A. If $\mathfrak{a}M = M$, then $M = 0$.

Proof We prove this by contradiction. Assume that $M \neq 0$ and let $x_1, \ldots, x_n$ constitute a *minimal* set of generators of M. Obviously, $n \geqslant 1$, since $M \neq 0$. Now $x_1 \in M = \mathfrak{a}M$ can be written as a linear combination $x_1 = a_1y_1 + \cdots + a_ry_r$ for some $a_1, \ldots, a_r \in \mathfrak{a}$ and $y_1, \ldots, y_r \in M$. Each y_i, on the other hand, is an A-linear combination of $x_1, \ldots, x_n$ and, therefore, since $\mathfrak{a}$ is an ideal, we can write $x_1 = b_1x_1 + \cdots + b_nx_n$ for some $b_1, \ldots, b_n \in \mathfrak{a}$. This can be rewritten as $(1 - b_1)x_1 = b_2x_2 + \cdots + b_nx_n$. But $b_1 \in \mathfrak{J}_A$ and hence by the characterization of elements of $\mathfrak{J}_A$ (See Exercise 1.2.11) $1 - b_1$ is a unit in A. Thus $x_1 = (1 - b_1)^{-1}b_2x_2 + \cdots + (1 - b_1)^{-1}b_nx_n$. If $n = 1$, then $x_1 = 0$, whereas if $n > 1$, then $x_1 \in \sum_{i=2}^{n} Ax_i$. In both these cases the minimality of the generating set $\{x_1, \ldots, x_n\}$ is contradicted. ◀

So far we have treated modules just as additive Abelian groups with scalar multiplication maps. The additive group of any ring R is an Abelian group. If we can give R an A-module structure (for some ring A) such that the multiplication of R is compatible with the scalar multiplication map $A \times R \to R$, R is called an algebra over A.

Let $\varphi : A \to R$ be a homomorphism of rings. Then the ring R possesses an A-module structure with the scalar multiplication map $ax := \varphi(a)x$ for $a \in A$ and $x \in R$. Furthermore, the ring structure and the A-module structure of R are compatible in the sense that for every $a, b \in A$ and $x, y \in R$ we have $(ax)(by) = (ab)(xy)$.

Conversely if a ring R has an A-module structure with $(ax)(by) = (ab)(xy)$ for every $a, b \in A$ and $x, y \in R$, then there is a unique ring homomorphism $\varphi : A \to R$ taking $a \mapsto a \cdot 1$ (where 1 denotes the identity of R and $\cdot$ denotes scalar multiplication). This motivates us to define the following.

1.39 Definition Let A be a ring. An algebra over A or an A-algebra is a ring R together with a ring homomorphism $\varphi : A \to R$. The homomorphism φ is called the structure homomorphism of the A-algebra R. If R and S are A-algebras with structure homomorphisms $\varphi : A \to R$ and $\psi : A \to S$, then an A-algebra homomorphism (from R to S) is a ring homomorphism $\eta : R \to S$ such that $\psi = \eta \circ \varphi$.

1.40 Example Let A be a ring.

(1) The polynomial ring $A[X_1, \ldots, X_n]$ (for indeterminates $X_1, \ldots, X_n$) is an A-algebra with the canonical inclusion as the structure homomorphism and is called a polynomial algebra over A.

(2) If $\mathfrak{a}$ is an ideal of A, then the canonical surjection $A \to A/\mathfrak{a}$ makes $A/\mathfrak{a}$ an A-algebra.

(3) If R is an A-algebra with the structure homomorphism $\varphi : A \to R$ and if S is an R-algebra with the structure homomorphism $\psi : R \to S$, then S is an A-algebra with the structure homomorphism $\psi \circ \varphi$.

(4) Combining (2) and (3) one can show that if R is an A-algebra and $\mathfrak{a}$ an ideal of R, then the ring $R/\mathfrak{a}$ is an A-algebra. This is called the quotient algebra of R by $\mathfrak{a}$.

Let $x_1, \ldots, x_n$ belong to an A-algebra R. Because of the A-module structure of R we can talk about the A-linear combinations of x_i. Now R being a ring, it also makes sense to talk about the products $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for non-negative integers α_i and to investigate what all these products generate as an A-module. This leads to the concept of algebra generators.

1.41 Definition Let R be an A-algebra with the structure homomorphism $\varphi : A \to R$. A subset S of R is said to generate R as an A-algebra, if every element $x \in R$ can be written as a polynomial expression in finitely many elements of S with coefficients from A (i.e., from $\varphi(A)$). We write this as $R = A[S]$.

If $S = \{x_1, \ldots, x_n\}$ is finite, we write $A[x_1, \ldots, x_n]$ in place of $A[S]$ and say that R is a finitely generated A-algebra and that the homomorphism $\varphi : A \to R$ is of finite type. On the other hand, if R is finitely generated as an A-module, then we say that R is a finite A-algebra.

It is important that the reader understands the distinction between the concepts of R generated as an A-algebra and R generated as an A-module. Since every A-linear combination is also a polynomial expression with coefficients from A, it follows that a generating set of R as an A-module is also a generating set of R as an A-algebra. In particular, finite A-algebras are also finitely generated A-algebras. The converse of this is, however, not true in general.

1.42 Example (1) The polynomial algebra $A[X_1, \ldots, X_n]$, $n \geqslant 1$, over A is not finitely generated as an A-module, but is finitely generated as an A-algebra. In fact, $A[X_1, \ldots, X_n]$ is a free A-module generated by the monomials $X_1^{\alpha_1} \cdots X_n^{\alpha_n}$ for all $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$.

(2) For an ideal $\mathfrak{a}$ of $A[X_1, \ldots, X_n]$ the ring $R := A[X_1, \ldots, X_n]/\mathfrak{a}$ is generated as an A-algebra by the equivalence classes $x_i := X_i + \mathfrak{a}$, $1 \leqslant i \leqslant n$. Thus we have $R = A[x_1, \ldots, x_n]$. If $\mathfrak{a}$ is not the zero ideal, then R is not, in general, (isomorphic to) a polynomial algebra over A. In fact $x_1, \ldots, x_n$ are not indeterminates (over A) in the sense that they satisfy non-zero polynomial equations $f(x_1, \ldots, x_n) = 0$ for every $f \in \mathfrak{a} \setminus \{0\}$. (In this case we also say that $x_1, \ldots, x_n$ are algebraically dependent.) In other words, the notation $A[\ldots]$ is a generalization of the notation to denote polynomial algebras. In what follows I will usually denote polynomial algebras by $A[X_1, \ldots, X_n]$ with upper-case letters as algebra generators, whereas for an arbitrary finitely generated A-algebra I use lower-case letters for the algebra generators as in $A[x_1, \ldots, x_n]$.

One may proceed to define kernels and images of A-algebra homomorphisms and frame and prove the isomorphism theorem for A-algebras. The details are left to the reader. Let me only mention here that algebra homomorphisms (and isomorphisms etc.) are essentially ring homomorphisms with the added condition of commutativity with the structure homomorphisms.

1.43 Theorem Let A be a ring. Then a ring R is a finitely generated A-algebra, if and only if R is a quotient of a polynomial algebra over A.

Proof [if] Immediate from Example 1.42.

[only if] Let $R = A[x_1, \ldots, x_n]$. Then the map $\eta : A[X_1, \ldots, X_n] \to R$ taking $f(X_1, \ldots, X_n) \mapsto f(x_1, \ldots, x_n)$ is a surjective A-algebra homomorphism. By the isomorphism theorem one then has the isomorphism $R \cong A[X_1, \ldots, X_n]/\operatorname{Ker}\eta$ of A-algebras. ◀

This theorem shows that for the study of finitely generated algebras it is sufficient to investigate the polynomial algebras and the quotients of the polynomial algebras.

Exercises for Section 1.3

1. Show that for every $n \in \mathbb{N}$ there are integers $a_1, \ldots, a_n$ that constitute a *minimal* set of generators for the unit ideal in $\mathbb{Z}$. (**Hint:** Take any n distinct primes $p_1, \ldots, p_n$. Define $a := \prod_{i=1}^n p_i$ and take $a_i := a/p_i$ for $i = 1, \ldots, n$.)

2. Let A be a ring. Prove or disprove:

*** (a)** Every A-submodule of a free A-module is again free.

(b) Every A-submodule of a non-free A-module is again non-free.

3. Prove Lemma 1.36.

4. Let M be a finitely generated A-module. Define

$$\mu_A(M) := \min\{|S| \mid M \text{ is generated by } S\}\,.$$

Show that if N is a submodule of M, then $\mu_A(M) \leqslant \mu_A(N) + \mu_A(M/N)$. Give an example where the strict inequality holds.

5. Let M be an A-module and N an A-submodule of M. Define

$$(M : N) := \{a \in A \mid aM \subseteq N\} \subseteq A\,.$$

(a) Show that $(M : N)$ is an ideal of A. In particular, for $N = 0$ the ideal $(M : 0)$ is called the (A-)annihilator of M and is denoted as $\operatorname{Ann}_A M$ (or as $\operatorname{Ann} M$).

(b) Let $\mathfrak{a} \subseteq \operatorname{Ann} M$ be an ideal of A. Show that M is an $A/\mathfrak{a}$-module under the scalar multiplication map $(a + \mathfrak{a})x := ax$. (**Remark:** It is only necessary to check that this map is well-defined, that is, the definition is independent of the choice of the representative a of the equivalence class $a + \mathfrak{a}$.)

6. Let M be an A-module. An element $x \in M$ is called a torsion element of M, if $\operatorname{Ann} Ax \neq 0$, that is, if there is an $a \in A \setminus \{0\}$ with $ax = 0$. The set of all torsion elements of M is denoted by $\operatorname{Tors} M$ (or $\operatorname{Tors}_A M$ if the ring A is to be highlighted). M is called torsion-free, if $\operatorname{Tors} M = 0$, and a torsion module, if $\operatorname{Tors} M = M$.

(a) Show that $\operatorname{Tors} M$ is a submodule of M.

(b) Show that $\operatorname{Tors} M$ is a torsion module (called the torsion submodule of M) and that $M/\operatorname{Tors} M$ is torsion-free.

(c) If A is an integral domain, show that every free module over A is torsion-free. In particular, every vector space is torsion-free.

7. Show that:

(a) $\mathbb{Q}$ is not finitely generated as a $\mathbb{Z}$-module. (**Hint:** If N is the $\mathbb{Z}$-submodule of $\mathbb{Q}$ generated by a_i/b_i, $i = 1, \ldots, n$, with $\gcd(a_i, b_i) = 1$, then for any prime p that does not divide $b_1 \cdots b_n$ we have $1/p \notin N$.)

(b) $\mathbb{Q}$ is not a free $\mathbb{Z}$-module. (**Hint:** Any two distinct elements of $\mathbb{Q}$ are linearly dependent over $\mathbb{Z}$.)

(c) $\mathbb{Q}$ is a torsion-free $\mathbb{Z}$-module. (**Remark:** This shows that the converse of Exercise 1.3.6(c) is not true in general, that is, for an integral domain A every torsion-free A-module need not be free. However, if A is a PID and if M is a finitely generated torsion-free A-module, then M is free. The proof of this last statement is not that easy.)

8. Let A be a non-zero ring and X, Y, Z indeterminates (over A). Demonstrate the following ring (actually A-algebra) isomorphisms:

(a) $A[X] \cong A[Y] \cong A[Z]$.

(b) $A[X, Y] \cong A[X][Y]$.

(c) $A[X, Y]/\langle X \rangle \cong A[Z]$.

(d) $A[X, Y]/\langle X, Y \rangle \cong A$.

(e) $A[X]/aA[X] \cong (A/aA)[X]$ for any $a \in A$.

(f) $A[X]/\mathfrak{a}^e \cong (A/\mathfrak{a})[X]$, where $\mathfrak{a}$ is an ideal of A and where $\mathfrak{a}^e$ is the extension of $\mathfrak{a}$ in $A[X]$.

(g) $A[X, Y]/\langle X - Y \rangle \cong A[Z]$.

*** (h)** $A[X, Y]/\langle aX - bY \rangle \cong A[Z]$, where A is a PID and $0 \neq a, b \in A \setminus A^*$ are relatively prime.

1.4 Field extensions

With groups, rings, modules etc. it is often useful to investigate a smaller structure (subgroup, subring or submodule) sitting inside a bigger one. With fields, however, the usual practice is the converse. That is, if we have a field that is lacking some desirable properties, we extend the field to get *superfields* (more commonly designated as field extensions) that possess those properties. For example, we get the field $\mathbb{R}$ as an extension of $\mathbb{Q}$ in an attempt to make it 'complete' in the sense that every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. Once completion is achieved, our journey does not stop, because we see that $\mathbb{R}$ is still not

big enough so that every polynomial with real coefficients will have a real root. So we adjoin the fictitious element $\mathrm{i} = \sqrt{-1}$ to $\mathbb{R}$ in order to get the field $\mathbb{C}$ of complex numbers. It turns out that $\mathbb{C}$ is both complete (in terms of convergence of Cauchy sequences) and algebraically closed (in the sense that every polynomial with complex coefficients has a complex root). Thus we should now keep ourselves rather happy with $\mathbb{C}$ and make no further attempts to extend $\mathbb{C}$, unless there is an (esoteric) need to do so.

In this section I will reserve the (Roman) letters F, K, L to designate fields. I start with some basic properties of roots of polynomials.

1.44 Definition Let $f(X) \in K[X]$. An element a in K (or in any extension of K) is said to be a root of f, if $f(a) = 0$.

1.45 Proposition Let $f(X) \in K[X]$ and $a \in K$. Then $f(X) = (X - a)q(X) + f(a)$ for some $q(X) \in K[X]$. In particular, a is a root of $f(X)$, if and only if $X - a$ divides $f(X)$ (in $K[X]$).

Proof Recall that $K[X]$ is an ED. Euclidean division of $f(X)$ by $X-a$ gives $f(X) = (X-a)q(X)+r(X)$ with $\deg r(X) < \deg(X - a) = 1$. Thus $r(X)$ is a constant polynomial. Let us denote $r(X)$ by $r \in K$. Substituting $X = a$ gives $f(a) = r$, whence the first result follows. The last statement is an immediate consequence of this. ◀

1.46 Proposition Let f be a non-zero polynomial of $K[X]$ with $d := \deg f$. Then f can have at most d roots in K.

Proof We proceed by induction on d. The result clearly holds for $d = 0$. So assume that $d \geqslant 1$ and that the proposition holds for all polynomials in $K[X]$ of degree $d - 1$. If f has no roots in K, we are done. So assume that f has a root, say, $a \in K$. By Proposition 1.45 we have $f(X) = (X - a)g(X)$ for some $g(X) \in K[X]$. Now $\deg g = d-1$ and so by the induction hypothesis g has at most $d-1$ roots in K. Since K is a field (and hence does not contain non-zero zero divisors), it follows that the roots of f are precisely a and the roots of g. This establishes the induction step. ◀

It is easy to see that Proposition 1.46 continues to remain valid, if K is any integral domain (not necessarily a field). However, if K is not an integral domain, the proposition does not necessarily hold. For example, if $ab = 0$ with $0 \neq a, b \in K$, $a \neq b$, then the polynomial $X^2 + (b - a)X$ has at least three roots, namely, 0, a and $a - b$.

For a field extension $K \subseteq L$ and for a polynomial $f \in K[X]$ we can talk about the roots of f in L, since $f \in L[X]$ too. Clearly all the roots of f in K are also roots of f in L. However, the converse is not true in general. For example, the only roots of $X^4 - 1$ in $\mathbb{R}$ are ± 1, whereas the roots of the same polynomial in $\mathbb{C}$ are $\pm 1, \pm \mathrm{i}$. Indeed we have the following important result.

1.47 Proposition Let $f \in K[X]$ be a non-constant polynomial. Then there exists a field extension L of K such that f has a root in L.

Proof If f has a root in K, taking $L = K$ proves the proposition. So we assume that f has no root in K (which implies that every irreducible factor of f has degree $\geqslant 2$). In principle we do not require f to be irreducible. But if we consider a non-constant factor g of f, irreducible over K, we see that the roots of g in any extension L of K are roots of f in L too. Thus we may replace f by g and assume without loss of generality that f itself is irreducible. We then construct the field extension[2] $L := K[X]/\langle f \rangle$ and denote the equivalence class of X in L by α. (One also writes x, $\bar{X}$ or $[X]$ to denote this equivalence class.) It is clear that $f(\alpha) = 0 \in L$, that is, α is a root of $f(X)$ in L. ◀

[2]Since $K[X]$ is a PID, $\langle f \rangle$ is a maximal ideal of $K[X]$ and hence L is indeed a field. Also K is canonically embedded in L.

We say that the field L in the proof of the last proposition is obtained by adjoining the root α of f and denote this as $L = K(\alpha)$. We write $f(X) = (X - \alpha)f_1(X)$, where $f_1(X) \in L[X]$ and $\deg f_1 = \deg f - 1$. Now there is a field extension L' of L, where f_1 has a root. Proceeding in this way one can prove the following result.

1.48 Proposition Let f be a non-constant polynomial in $K[X]$ with $\deg f = d$. Then there exists a field extension L of K such that f has d roots (not necessarily all distinct) in L.

If a polynomial $f \in K[X]$ of degree $d \geqslant 1$ has all its roots $\alpha_1, \ldots, \alpha_d$ in L, then we have the factorization $f(X) = a(X - \alpha_1) \cdots (X - \alpha_d)$ for some $a \in L$ (actually $a \in K$). In this case we say that f splits (completely or into linear factors) over L.

1.49 Definition Let $f \in K[X]$ be a non-constant polynomial. A minimal (with respect to inclusion) field extension of K over which f splits completely is called a splitting field of f over K. This is a minimal field which contains K and all the roots of f.

From the above discussion it is clear that every non-constant polynomial $f \in K[X]$ has a splitting field L. Quite importantly the field L is unique in some sense. This allows us to call *the* splitting field of f instead of *a* splitting field of f. I will discuss these topics again later. For the time being let me mention that the phrase 'over K' is necessary in the definition of splitting fields. For example, the splitting field of $X^2 + 1$ over $\mathbb{Q}$ is not the same as that of the same polynomial over $\mathbb{R}$.

1.50 Definition Let f be a non-constant polynomial in $K[X]$ and let α be a root of f (in some extension of K). The largest natural number n for which $(X - \alpha)^n \mid f(X)$ is called the multiplicity of the root α (in f). If $n = 1$ (resp. $n > 1$), then α is called a simple (resp. multiple) root of f. If all the roots of f are simple, then we call f a square-free polynomial. It is easy to see that f is square-free, only if f is not divisible by the square of a non-constant polynomial in $K[X]$. The reverse implication also holds, if $\operatorname{char} K = 0$ or if K is a finite field.

The notion of multiplicity can be extended to a non-root β of f by setting the multiplicity of β to zero.

Now for a while let us assume that $K \subseteq L$ is a field extension.

1.51 Definition An element $\alpha \in L$ is said to be algebraic over K, if there exists a non-constant polynomial $f(X) \in K[X]$ with $f(\alpha) = 0$. If an element $\alpha \in L$ is not algebraic over K, then we say that α is transcendental over K. Thus a transcendental (over K) element $\alpha \in L$ is a root of *no* polynomial in $K[X]$. The field extension $K \subseteq L$ is called an algebraic extension, if every element of L is algebraic over K. A non-algebraic extension is also often called a transcendental extension. If $K \subseteq L$ is a transcendental extension, there exists at least one element $\alpha \in L$ which is transcendental (i.e., not algebraic) over K.

1.52 Example (1) Every element $\alpha \in K$ is algebraic over K, since it is a root of the non-constant polynomial $X - \alpha \in K[X]$.

(2) The element $\alpha := \sqrt[3]{2 + \sqrt{3}} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$, since α is a root of the polynomial $(X^3 - 2)^2 - 3 = X^6 - 4X^3 + 1 \in \mathbb{Q}[X]$.

(3) The well-known real numbers e and π are transcendental over $\mathbb{Q}$. (We are not going to prove this.) Note that the concepts of algebraic and transcendental elements are heavily dependent on the field K. For example, e and π, being elements of $\mathbb{R}$, are algebraic over $\mathbb{R}$.

(4) Let $z := a + \mathrm{i}b \in \mathbb{C}$, where $\mathrm{i} := \sqrt{-1}$ and $a, b \in \mathbb{R}$. Then z is a root of the polynomial $(X-a)^2 + b^2 = X^2 - 2aX + (a^2+b^2) \in \mathbb{R}[X]$. It, therefore, follows that every complex number is algebraic over $\mathbb{R}$. In other words, the field extension $\mathbb{R} \subseteq \mathbb{C}$ is algebraic.

(5) The extension $\mathbb{Q} \subseteq \mathbb{R}$ is transcendental, since $\mathbb{R}$ contains elements (like e and π) that are transcendental over $\mathbb{Q}$. Therefore, the extension $\mathbb{Q} \subseteq \mathbb{C}$ is also transcendental.

1.53 Definition Let $\alpha \in L$ be algebraic over K. A non-constant polynomial $f \in K[X]$ of least (positive) degree with $f(\alpha) = 0$ is called a $\mathrm{minimal\ polynomial}$ of α over K.

1.54 Proposition Let $\alpha \in L$ be algebraic over K. A minimal polynomial f of α over K is irreducible over K. If $h \in K[X]$ is a polynomial with $h(\alpha) = 0$, then $f \mid h$ in $K[X]$. In particular, any two minimal polynomials f and g of α satisfy $g(X) = cf(X)$ for some $c \in K^*$.

Proof If f is reducible over K, then $f = f_1 f_2$ for some non-constant polynomials $f_1, f_2 \in K[X]$. Since K is a field and $0 = f(\alpha) = f_1(\alpha) f_2(\alpha)$, we have $f_1(\alpha) = 0$ or $f_2(\alpha) = 0$. But $1 \leqslant \deg f_1 < \deg f$ and $1 \leqslant \deg f_2 < \deg f$, a contradiction to the choice of f.

Using polynomial division one can write $h(X) = q(X)f(X) + r(X)$ for some polynomials $q, r \in K[X]$. Now $h(\alpha) = 0$ implies $r(\alpha) = 0$. Since $\deg r < \deg f$, the choice of f forces $r(X) = 0$, i.e., $f \mid h$.

Finally if f and g are two minimal polynomials of α over K, then $f \mid g$ and $g \mid f$, i.e., $g(X) = cf(X)$ for some unit c of $K[X]$. But the units of $K[X]$ are precisely the non-zero elements of K. ◀

If f is a monic minimal polynomial of α over K, then by the last proposition f is uniquely determined by α and K. It is, therefore, customary to define *the* minimal polynomial of α over K to be this (unique) monic polynomial. Unless otherwise stated we will also stick to this revised definition and use the symbol $\mathrm{minpoly}_{\alpha,K}(X) \in K[X]$ to denote the minimal polynomial of α over K. If K is clear from the context, we may simply write $\mathrm{minpoly}_{\alpha}(X)$.

1.55 Example (1) The minimal polynomial of $\alpha \in K$ over K is the linear polynomial $X - \alpha \in K[X]$.

(2) A complex number $z = a + \mathrm{i}b$, $a, b \in \mathbb{R}$, $b \neq 0$, is not a root of a linear polynomial over $\mathbb{R}$. On the other hand, z is a root of the quadratic polynomial $f(X) = X^2 - 2aX + (a^2+b^2) \in \mathbb{R}[X]$. It follows that f is the minimal polynomial of z over $\mathbb{R}$ and thus f is irreducible in $\mathbb{R}[X]$.

1.56 Proposition For a field K the following conditions are equivalent:

(a) Every proper field extension $K \subsetneqq L$ is transcendental, i.e., K has no algebraic extension other than itself.

(b) Every non-constant polynomial in $K[X]$ has a root in K.

(c) Every non-constant polynomial in $K[X]$ splits in K.

(d) Every non-constant irreducible polynomial in $K[X]$ is of degree 1.

Proof [(a)⇒(b)] Consider a non-constant irreducible polynomial $f(X) \in K[X]$ and the field extension $L := K[X]/\langle f \rangle$ of K. We have seen that L contains a root of f. We will prove later (Corollary 1.64) that this extension $K \subseteq L$ is algebraic. Hence (a) implies that $L = K$, that is, K contains a root of f.

[(b)⇒(c)] Let $f \in K[X]$ be a non-constant polynomial. By (b) f has a root, say, α_1 in K. Thus $f(X) = (X - \alpha_1) f_1(X)$ for some $f_1 \in K[X]$ with $\deg f_1 = \deg(f) - 1$. If f_1 is a constant polynomial, we are done. Otherwise, we find as above $\alpha_2 \in K$ and $f_2 \in K[X]$ with $f_1(X) = (X - \alpha_2) f_2(X)$ and with $\deg f_2 = \deg(f_1) - 1 = \deg(f) - 2$. Proceeding in this way proves (c).

[(c)⇒(d)] Obvious.

[(d)⇒(a)] Let $K \subseteq L$ be an algebraic extension, $\alpha \in L$ and $f(X) := \mathrm{minpoly}_{\alpha,K}(X) \in K[X]$. Since f is irreducible over K, by (d) we have $\deg f = 1$, i.e., $f(X) = X - \alpha$, i.e., $\alpha \in K$. Thus $L \subseteq K$, i.e., $L = K$. ◀

1.57 Definition A field K satisfying one (and hence all) of the equivalent conditions of Proposition 1.56 is called an algebraically closed field. For any arbitrary field K a minimal algebraically closed field containing K is called an algebraic closure of K and is denoted by $\bar{K}$. If L is an algebraically closed field containing K, there exists a field K' with $K \subseteq K' \subseteq L$ such that K' is an algebraic closure of K. We call K' an algebraic closure of K in L.

We will see soon that an algebraic closure of every field exists and is unique in some sense. The algebraic closure of an algebraically closed field K is K itself. The following is a very well-known result. I will not prove the theorem here. Interested students may consult Exercise 1.4.1.

1.58 Theorem [Fundamental theorem of algebra] The field $\mathbb{C}$ of complex numbers is algebraically closed.

$\mathbb{R}$ is not algebraically closed, since the proper extension $\mathbb{C}$ of $\mathbb{R}$ is algebraic (Example 1.52). Indeed $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$. The algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$ is a proper subfield of $\mathbb{C}$ (Exercise 1.4.8(d)).

I now introduce an important quantity associated with a field extension. Recall that if $F \subseteq K$ is a field extension, then K is a vector space over F.

1.59 Definition For a field extension $F \subseteq K$ the cardinality of any F-basis of K is called the degree of extension or the extension degree of K over F and is usually denoted by $[K : F]$. Thus $[K : F] = \dim_F K$. If $[K : F]$ is finite, we say that K is a finite extension of F. Otherwise, the extension is said to be infinite.

1.60 Example Let $f(X) \in F[X]$ be irreducible (over F) of degree $d \geqslant 1$. Then $K := F[X]/\langle f(X)\rangle$ is a field extension of F. One can easily check that the equivalence classes of $1, X, \ldots, X^{d-1}$ form an F-basis of K. Thus $[K : F] = d$.

1.61 Proposition Let $F \subseteq K \subseteq L$ be a tower of field extensions. Then $[L : F] = [L : K][K : F]$. In particular, the extension $F \subseteq L$ is finite, if and only if the extensions $F \subseteq K$ and $K \subseteq L$ are finite. Furthermore, if $[L : F]$ is finite, then $[L : K] \mid [L : F]$ and $[K : F] \mid [L : F]$.

Proof For an F-basis S of K and a K-basis S' of L consider the set $T := \{xy \mid x \in S \text{ and } y \in S'\} \subseteq L$. It can be easily verified that T generates L as an F-vector space and that T is linearly independent over F. The details are left to the reader. ◀

Let $F \subseteq K$ be a field extension and $a \in K$. Then we define

$$F[a] := \{f(a) \mid f(X) \in F[X]\}$$

and

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

It is easy to see that $F[a]$ is the smallest (with respect to inclusion) of the integral domains (contained in K) that contain F and a. On the other hand, $F(a)$ is the smallest of the fields (contained in K) that contain

F and a. We also have $F[a] \subseteq F(a)$. Now we prove the following important characterization of algebraic elements.

1.62 Theorem For a field extension $F \subseteq K$ and an element $a \in K$ the following conditions are equivalent:

(a) The element a is algebraic over F.

(b) The extension $F(a)$ is finite over F (i.e., $[F(a) : F] < \infty$).

(c) $F(a) = F[a]$.

Proof [(a)⇒(b)] Let $h(X) := \mathrm{minpoly}_{a,F}(X) \in F[X]$ and $d := \deg h$. Consider the ring homomorphism $\varphi : F[X] \to F(a)$ that takes $f(X) \mapsto f(a)$. It follows from Proposition 1.54 that $\mathrm{Ker}\,\varphi = \langle h \rangle$. Therefore, by the isomorphism theorem we have $F[X]/\langle h \rangle \cong \mathrm{Im}\,\varphi$. Since h is irreducible over F, $F[X]/\langle h \rangle$ is a field of extension degree d over F. Therefore, we are done, if we can show that $\mathrm{Im}\,\varphi = F(a)$. But since $\mathrm{Im}\,\varphi$ is a field containing F and a (Note that $\varphi(X) = a$.), it is immediate that $F(a) \subseteq \mathrm{Im}\,\varphi$, that is, $\mathrm{Im}\,\varphi = F(a)$.

[(b)⇒(c)] Let $d := [F(a) : F]$. Since the $d + 1$ elements $1, a, a^2, \ldots, a^d$ are linearly dependent over F, there exist $\alpha_0, \ldots, \alpha_d \in F$, not all 0, such that $\alpha_0 + \alpha_1 a + \cdots + \alpha_d a^d = 0$. This, in turn, implies that there is an irreducible polynomial $h(X) \in F[X]$ with $h(a) = 0$. Now consider $f(a)/g(a) \in F(a)$. Clearly $h \not| g$ (because otherwise $g(a) = 0$). Since h is irreducible, $\gcd(g, h) = 1$, i.e., there exist polynomials $u(X), v(X) \in F[X]$ with $u(X)g(X) + v(X)h(X) = 1$, i.e., with $u(a)g(a) = 1$. But then $f(a)/g(a) = u(a)f(a) \in F[a]$. Thus $F(a) \subseteq F[a]$. The reverse inclusion is obvious.

[(c)⇒(a)] Clearly the element 0 is algebraic over F. So assume $a \neq 0$. Since $1/a \in F(a)$, by hypothesis there is a polynomial $f(X) \in F[X]$ such that $1/a = f(a)$. But then a is a root of the non-constant polynomial $Xf(X) - 1 \in F[X]$. ◀

1.63 Corollary Let $F \subseteq K$ be a field extension. Then the set of elements in K that are algebraic over F is a field.

Proof It is sufficient to show that if $a, b \in K$ are algebraic over F, then the elements $a \pm b$, ab and a/b (if $b \neq 0$) are also algebraic over F. By the last theorem $[F(a) : F]$ is finite. Since b is algebraic over F, it is also algebraic over $F(a)$. In particular, $[F(a)(b) : F(a)]$ is finite. But then by Proposition 1.61 the extension $F(a)(b)$ is finite over F and contains $a \pm b$, ab and a/b (if $b \neq 0$). ◀

The field $F(a)(b)$ in the proof of the last corollary is also denoted as $F(a, b)$. It is, in fact, the smallest subfield of K that contains F, a and b, and it follows that $F(a, b) = F(b, a)$. More generally, for a field extension $F \subseteq K$ and for elements $a_1, \ldots, a_n \in K$ each algebraic over F the field $F(a_1, \ldots, a_n)$ is defined as $F(a_1)(a_2)\cdots(a_n)$ and is independent of the order in which a_i are adjoined.

1.64 Corollary Let $F \subseteq K$ be a finite extension. Then K is algebraic over F.

Proof For any $a \in K$ the degree $[F(a) : F]$ divides $[K : F]$ and hence is finite. ◀

The converse of the last corollary is, however, not true. That is, it is possible that an algebraic extension of a field F has infinite extension degree over F. (See Exercise 1.4.6 as an example.)

1.65 Corollary If $F \subseteq K$ and $K \subseteq L$ are algebraic field extensions, then $F \subseteq L$ is also algebraic.

Proof Let $a \in L$. $K \subseteq L$ being algebraic, there exists $f(X) := \alpha_n X^n + \alpha_{n-1} X^{n-1} + \cdots + \alpha_0 \in K[X]$ with $f(a) = 0$. It then follows that a is algebraic over $F(\alpha_0, \ldots, \alpha_n)$. Since each α_i is algebraic over F, $[F(\alpha_0, \ldots, \alpha_n) : F]$ is finite, so that

$$[F(\alpha_0, \ldots, \alpha_n)(a) : F] = [F(\alpha_0, \ldots, \alpha_n)(a) : F(\alpha_0, \ldots, \alpha_n)][F(\alpha_0, \ldots, \alpha_n) : F]$$

is also finite and hence the extension $F(\alpha_0, \ldots, \alpha_n)(a)$ of F and, in particular, a is algebraic over F. ◀

1.66 Definition A field extension $F \subseteq K$ is called simple, if $K = F(a)$ for some $a \in K$. In this case a is called a primitive element of K (over F).

1.67 Proposition Let F be a field of characteristic 0 and let the elements a, b (belonging to some extension of F) be algebraic over F. Then the extension $F(a, b)$ is simple.

Proof Let $p(X)$ and $q(X)$ be the respective minimal polynomials of a and b over F. Let $d := \deg p$ and $d' := \deg q$. The polynomials p and q are irreducible over F and hence by Exercise 1.4.3 have no multiple roots. Let $a_1, \ldots, a_d$ be the roots of p and $b_1, \ldots, b_{d'}$ the roots of q with $a = a_1$ and $b = b_1$. For each i, j with $j \neq 1$ the equation $a_i + \lambda b_j = a + \lambda b$ has a unique solution for λ. Since F is infinite, we can choose $\mu \in F$ which is a solution of neither of the equations just mentioned. Define $c := a + \mu b$, so that $c \neq a_i + \mu b_j$ for all i, j with $j \neq 1$. Clearly $F(c) \subseteq F(a, b)$. To prove the reverse inclusion, note that by hypothesis $q(b) = 0$. Also if we define $f(X) := p(c - \mu X) \in F(c)[X]$, we see that $f(b) = p(a) = 0$. By choice of c we see that $f(b_j) \neq 0$ for $j \neq 1$. Finally since q is square-free, we have $\gcd(q, f) = X - b \in F(c)[X]$. This implies that $b \in F(c)$ and so $a = c - \mu b \in F(c)$. ◀

1.68 Corollary A finite extension $F \subseteq K$ of fields of characteristic 0 is simple.

Proof We proceed by induction on $d := [K : F]$. The result vacuously holds for $d = 1$. So let's assume that $d > 1$ and that the result holds for all extensions of degree $< d$ of fields of characteristic 0. Choose $a \in K \setminus F$. Then $[F(a) : F] > 1$, so that $[K : F(a)] = [K : F]/[F(a) : F] < d$. By the induction hypothesis the extension $F(a) \subseteq K$ is simple, say $K = F(a)(b) = F(a, b)$. The result now follows immediately from the previous proposition. ◀

Now we have sufficient machineries to prove the existence and uniqueness of splitting fields of polynomials. Let f be an arbitrary non-constant polynomial of degree d in $F[X]$. Assume that f does not split over F and consider an irreducible factor f_1 of f of degree $d_1 > 1$. $F_1 := F[X]/\langle f_1 \rangle$ is a field extension of F. If α_1 denotes the equivalence class of X in F_1, then the elements $1, \alpha_1, \ldots, \alpha_1^{d_1 - 1}$ constitute a basis of F_1 over F. In particular, $[F_1 : F] = d_1 \leqslant d$. Now one can write $f(Y) = (Y - \alpha_1)g(Y)$ for some $g(Y) \in F_1[Y]$. If g splits over F_1, then f also does so. Otherwise, choose any irreducible factor g_1 of g with $\deg g_1 > 1$ and consider the field extension $F_2 := F_1[Y]/\langle g_1 \rangle$ of F_1. Then $[F_2 : F_1] = \deg g_1 \leqslant \deg g = d - 1$, so that $[F_2 : F] \leqslant d(d-1)$. Moreover, if α_2 is the equivalence class of Y in F_2, then $f(Z) = (Z - \alpha_1)(Z - \alpha_2)h(Z)$ for some polynomial $h(Z) \in F_2[Z]$. Proceeding in this way we can prove the following result.

1.69 Proposition For a polynomial $f \in F[X]$ of degree $d \geqslant 1$ there is a field extension K of F with $[K : F] \leqslant d!$ such that f splits over K. ◀

That's the existence of splitting fields. Now comes the question of uniqueness. Let $\mu : F \to F'$ be an isomorphism of fields. Then μ induces an isomorphism $\mu^* : F[X] \to F'[Y]$ of polynomial rings defined by $a_d X^d + a_{d-1} X^{d-1} + \cdots + a_0 \mapsto \mu(a_d) Y^d + \mu(a_{d-1}) Y^{d-1} + \cdots + \mu(a_0)$. Note that $\mu^*(a) = \mu(a)$ for all $a \in F$. We also see that $f \in F[X]$ is irreducible over F, if and only if $\mu^*(f) \in F'[Y]$ is irreducible over F'. With these notations we state the following important lemma.

1.70 Lemma Let the non-constant polynomial $f \in F[X]$ be irreducible over F. Let α and β be any roots of f and $\mu^*(f)$ respectively. Then there is an isomorphism $\nu : F(\alpha) \to F'(\beta)$ of fields, such that $\nu(a) = \mu(a)$ for all $a \in F$ and $\nu(\alpha) = \beta$.

Proof Since $F(\alpha) = F[\alpha]$ and $F'(\beta) = F'[\beta]$, we define the map $\nu : F[\alpha] \to F'[\beta]$ by $g(\alpha) \mapsto (\mu^*(g))(\beta)$ for each $g \in F[X]$. It is now an easy check that ν is a well-defined isomorphism of fields with the desired properties. ◀

Roots of an irreducible polynomial are called conjugates of one another. If α and β are two roots of a non-constant irreducible polynomial $f(X) \in F[X]$, then the last lemma guarantees the existence of an isomorphism $\tau : F(\alpha) \to F(\beta)$ that fixes all the elements of F and maps $\alpha \mapsto \beta$.

1.71 Proposition We use the maps μ and μ^* as defined above. Let $f(X) \in F[X]$ be a non-constant polynomial and let K and K' be some splitting fields of f and $\mu^*(f)$ (over F and F') respectively. Then there is an isomorphism $\tau : K \to K'$ of fields, such that for all $a \in F$ we have $\tau(a) = \mu(a)$.

Proof We proceed by induction on $d := [K : F]$. (By Proposition 1.69 d is finite.) If $d = 1$, the polynomial f splits over F itself and since K is a minimal field containing F and all the roots of f, we must have $K = F$. It also follows that $\mu^*(f)$ splits over F' and hence $K' = F'$. Thus $\tau = \mu$ is the desired isomorphism.

Now assume that $d > 1$ and that the result holds for all fields L and for all polynomials in $L[X]$ with splitting fields (over L) of extension degrees less than d. Consider an irreducible factor g of f with $1 < \deg g \leqslant \deg f$. Note that g also splits over K. We take any root $\alpha \in K$ of g and consider the intermediate field $F(\alpha)$, i.e., $F \subseteq F(\alpha) \subseteq K$. Similarly, let $\beta \in K'$ be a root of $\mu^*(g)$ and we consider the tower of extensions $F' \subseteq F'(\beta) \subseteq K'$. By Lemma 1.70 there is an isomorphism $\nu : F(\alpha) \to F'(\beta)$ with $\nu(a) = \mu(a)$ for all $a \in F$ and $\nu(\alpha) = \beta$. One can extend ν to $\nu^* : F(\alpha)[X] \to F'(\beta)[Y]$ as before. We then clearly have $\nu^*(f) = \mu^*(f)$. Now $[K : F(\alpha)] = [K : F]/[F(\alpha) : F] = [K : F]/\deg g < d$. It is evident that K and K' are splitting fields of f and $\nu^*(f)$ over $F(\alpha)$ and $F'(\beta)$ respectively. Hence by the induction hypothesis there is an isomorphism $\tau : K \to K'$ with $\tau(a) = \nu(a)$ for all $a \in F(\alpha)$. In particular, $\tau(a) = \mu(a)$ for all $a \in F$. ◀

The results pertaining to the splitting field of a polynomial can be generalized in the following way. Let S be a set of non-constant polynomials of $F[X]$. Then a splitting field of S over F is a minimal field K containing F over which each polynomial $f \in S$ splits. If $S = \{f_1, \ldots, f_r\}$ is a finite set, then the splitting field of S is the same as the splitting field of $f = f_1 \cdots f_r$ (See Exercise 1.4.4). But the situation is different, if S is infinite. Of particular interest to us is the set S consisting of all non-constant irreducible polynomials in $F[X]$. In this case the splitting field of S is an algebraic closure of F.

We give a sketch of the proof that even when S is infinite, a splitting field for S exists. This, in particular, establishes the existence of an algebraic closure of any field. For each $f \in S$ we define an indeterminate X_f and consider the ring $A := F[X_f \mid f \in S]$ and the ideal $\mathfrak{a}$ generated by $f(X_f)$ for all $f \in S$. We have $\mathfrak{a} \neq A$ and, therefore, there is a maximal ideal $\mathfrak{m}$ of A containing $\mathfrak{a}$ (Proposition 1.26). Consider the field $F_1 := A/\mathfrak{m}$ containing F. It follows that every polynomial $f \in S$ contains at least one root in F_1. Now we repeat the above procedure with F replaced by F_1 and S replaced by the set S_1 of all non-constant irreducible (over F_1) factors of polynomials in S to get another field F_2 containing F_1 (and hence F). We continue this procedure (infinitely often, if necessary) getting a sequence of fields $F \subseteq F_1 \subseteq F_2 \subseteq F_3 \subseteq \cdots$ and define K to be the field consisting of all elements of $\bigcup_{n\in\mathbb{N}} F_n$ that are algebraic over F. Then each polynomial in S splits in K, but in no proper subfield of K. So K is a splitting field of S.

Now let S be the set of all non-constant irreducible polynomials of $F[X]$. We want to show that the field K obtained as above is algebraically closed in this case. Let $K \subseteq L$ be an algebraic extension. Since the extensions $F \subseteq K$ is also algebraic, so is the extension $F \subseteq L$. Take any $\alpha \in L$. Then $h(X) := \mathrm{minpoly}_{\alpha,F}(X) \in F[X]$ is irreducible over F and by the construction of K has all the roots in K. In particular, $\alpha \in K$, i.e., $L = K$.

It is also true that the splitting field of S is unique upto isomorphisms that fix elements of F. In particular, the algebraic closure of F is unique upto isomorphisms that fix elements of F. We are not going to prove this uniqueness here.

Exercises for Section 1.4

1. Let $f(z)$ be a non-constant polynomial with complex coefficients.

(a) Show that $f(z)$ is unbounded (i.e., $|f(z)| \to \infty$ as $|z| \to \infty$).

(b) Show that $f(z)$ has a root $z_0 \in \mathbb{C}$. (**Hint:** Assume not, i.e., $f(z) \neq 0$ for all $z \in \mathbb{C}$. Then $g(z) := 1/f(z)$ is a bounded entire function and hence by Liouville's theorem is a constant.)

2. Show that the irreducible polynomials in $\mathbb{R}[X]$ have degrees $\leqslant 2$. (**Hint:** Use the fundamental theorem of algebra.)

3. Let K be a field and $f(X) = a_n X^n + \cdots + a_1 X + a_0 \in K[X]$. The formal derivative f' of f is defined to be the polynomial $f'(X) := \sum_{j=1}^{n} j a_j X^{j-1} \in K[X]$.

(a) Let $f, g \in K[X]$. Show that $(f+g)' = f' + g'$ and $(fg)' = f'g + fg'$.

(b) If $\operatorname{char} K = 0$, show that $f' = 0$, if and only if $f \in K$.

(c) If $\operatorname{char} K = p > 0$, then $f' = 0$, if and only if $f(X) = g(X^p)$ for some polynomial $g(X) \in K[X]$.

(d) Show that f ($\neq 0$) has no multiple roots (in any extension field of K), i.e., f is square-free, if and only if $\gcd(f, f') = 1$.

(e) Let f be a non-constant irreducible polynomial over K. Show that if $\operatorname{char} K = 0$, then f has no multiple roots. On the other hand, if $\operatorname{char} K = p > 0$, show that f has multiple roots, if and only if $f(X) = g(X^p)$ for some $g(X) \in K[X]$. (However, if $K = \mathbb{Z}_p$, then by Fermat's little theorem and by the binomial theorem $g(X^p) = g(X)^p$, which contradicts the fact that $f(x)$ is irreducible. Therefore, f cannot have multiple roots.)

4. Let $K \subseteq L$ be a field extension and $f_1, \ldots, f_n$ non-constant polynomials in $K[X]$. Show that each f_i, $i = 1, \ldots, n$, splits over L, if and only if the product $f_1 \cdots f_n$ splits over L.

5. Let $f(X) \in K[X]$ be irreducible of degree $d > 2$ and L the splitting field of f over K. Give an example when $[L : K] = d!$ and an example when $[L : K] < d!$.

6. Show that a finite field (i.e., a field with finite cardinality) is not algebraically closed. In particular, the algebraic closure of a finite field is infinite. (**Hint:** Let $a_1, \ldots, a_n$ be all the elements of a finite field K. Consider the polynomial $(X - a_1) \cdots (X - a_n) + 1 \in K[X]$.)

7. Let L be an algebraic closure of a field K. Prove that L is an algebraic extension of K.

8. A complex number z is called an algebraic number, if z is algebraic over $\mathbb{Q}$. An algebraic number z is called an algebraic integer, if z is a root of a *monic* polynomial in $\mathbb{Z}[X]$. Show that:

(a) If z is an algebraic number, then mz is an algebraic integer for some $m \in \mathbb{N}$.

(b) If $a \in \mathbb{Q}$ is an algebraic integer, then $a \in \mathbb{Z}$.

(c) If $z \in \mathbb{C}$ is an algebraic integer, then for any integer $n \in \mathbb{Z}$ the complex numbers nz and $z + n$ are algebraic integers.

(d) The set of all algebraic numbers is countable (and infinite). (**Remark:** This implies that the algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$ is countable. On the other hand, $\mathbb{C}$ is uncountable. Therefore, $\bar{\mathbb{Q}} \subsetneq \mathbb{C}$. This last statement also follows from Exercise 1.4.7 and Example 1.52(5).)

9. Let $f(X) \in K[X]$ be a non-constant polynomial of degree d and let $\alpha_1, \ldots, \alpha_d$ be the roots of f (in some extension field of K). The quantity $\Delta(f) := \prod_{1 \leqslant i < j \leqslant d} (\alpha_i - \alpha_j)^2$ is called the discriminant of f. Prove the following assertions:

(a) $\Delta(f) = 0$, if and only if f has a multiple root.

(b) $\Delta(f) \in K$.

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

(d) $\Delta(X^3 + aX + b) = -(4a^3 + 27b^2)$.

10. Let $F \subseteq K$ be a field extension and let φ be an endomorphism of K with $\varphi(a) = a$ for every $a \in F$.

(a) If an irreducible polynomial $f(X) \in F[X]$ has a root $\alpha \in K$, show that $\varphi(\alpha) \in K$ is also a root of f. For example, taking $F = \mathbb{R}$, $K = \mathbb{C}$ and φ the automorphism mapping z to its (complex) conjugate $\bar{z}$ allows us to conclude that if a complex number z is a root of the polynomial $f(X) \in \mathbb{R}[X]$, then $\bar{z}$ is also a root of f. A similar result holds for the extension $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{m})$, where m is a non-square rational number.

(b) If K is algebraic over F, show that φ is an automorphism. (**Hint:** Let the conjugates of $\alpha \in K$ over F be $\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_n$. Since φ is injective, it follows from Part (a) that φ makes a permutation of $\alpha_1, \ldots, \alpha_n$. Thus φ is surjective.)

11. Prove the following assertions:

(a) $\mathbb{R}$ is an infinite extension of $\mathbb{Q}$. (**Hint:** Consider transcendental numbers.)

(b) The only automorphism of $\mathbb{R}$ that fixes all the elements of $\mathbb{Q}$ is the identity map. (**Hint:** Let φ be such an automorphism. First note that if $0 \leqslant a \in \mathbb{R}$, then $\varphi(a) = \varphi(\sqrt{a})^2 \geqslant 0$. This implies that for $a, b \in \mathbb{R}$ with $a \leqslant b$ one has $\varphi(a) \leqslant \varphi(b)$. Now assume $a < \varphi(a)$ for some $a \in \mathbb{R} \setminus \mathbb{Q}$. Choose a rational number b with $a < b < \varphi(a)$. Then $\varphi(a) \leqslant \varphi(b) = b$, a contradiction. Thus $\varphi(a) \geqslant a$. Similarly $\varphi(a) \leqslant a$.)

1.5 Finite fields

A finite field is a field containing only finitely many elements. Though infinite fields like $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ are more familiar to us, finite fields often play important roles in algebra and number theory. The simplest example of a finite field is the field of residue classes of $\mathbb{Z}$ modulo a prime number p. Such a field, which we denote as $\mathbb{Z}_p$, consists of exactly p elements $[0]_p, [1]_p, \ldots, [p-1]_p$. But there are other finite fields too. Though they have rich algebraic structures, they are not as easy to visualize as the fields $\mathbb{Z}_p$. In this section we let p be a prime number and q a power of a prime, i.e., $q = p^n$ for some $n \in \mathbb{N}$. We will soon see that there exists a finite field with q elements. If $n \geqslant 2$, this field is not the same as the ring $\mathbb{Z}_q$. In fact, if q is composite, then $\mathbb{Z}_q$ has non-zero zero divisors and is not even an integral domain.

Recall that the characteristic of a ring A is the smallest *positive* integer n such that the sum $1+1+\cdots+1$ (n times) in A is the zero element of A. We denote this by $\operatorname{char} A = n$. If no such n exists, we take $\operatorname{char} A = 0$. A field K of characteristic zero (like $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$) has to be infinite, since 0, 1, $2 = 1 + 1$, $3 = 1 + 1 + 1, \ldots$ are distinct elements of K. Thus a finite field must have positive characteristic. In fact, if K is a finite field, $\operatorname{char} K$ has to be a prime. More generally, we have:

1.72 Proposition Let A be an integral domain of positive characteristic p. Then p is a prime.

Proof If p is composite, write $p = mn$ for some $m, n \in \mathbb{N}$, $1 < m < p$ and $1 < n < p$. But then $p = mn = 0$ (in A). Since A is an integral domain, we must have $m = 0$ or $n = 0$ (in A). This contradicts the minimality of p. ◀

Let K be a finite field of cardinality q and let $p := \operatorname{char} K \in \mathbb{P}$. K contains an isomorphic copy of the field $F := \mathbb{Z}_p$. If $[K : F] = n \in \mathbb{N}$, it follows that $q = p^n$ (since K is a $\mathbb{Z}_p$-vector space of dimension n). Therefore we have proved the first statement of the following important result.

1.73 Theorem The cardinality q of a finite field is a power p^n, $n \in \mathbb{N}$, of a prime number p. Conversely, given $p \in \mathbb{P}$ and $n \in \mathbb{N}$, there exists a finite field of cardinality $q = p^n$.

Proof In order to construct a finite field of cardinality $q = p^n$, we start with the field $F := \mathbb{Z}_p$ and consider the polynomial $f(X) := X^q - X \in \mathbb{Z}_p[X]$. Let K be *the* splitting field of f over F. Since $f'(X) = -1 \neq 0$, the roots of f are distinct (See Exercise 1.4.3). Therefore, the set $E := \{a \in K \mid a^q = a\}$ has cardinality q.

From Exercise 1.5.1 it follows that E is a field. But then $F \subseteq E \subseteq K$ and f splits over E. By the definition of splitting fields we must then have $K = E$, that is, $|K| = |E| = q$. ◀

1.74 Theorem Let K be a finite field of cardinality $q = p^n$ and let F be the subfield of K isomorphic to $\mathbb{Z}_p$. Then K is the splitting field of the polynomial $f(X) := X^q - X \in F[X]$ over F. In particular, K is unique up to isomorphisms fixing elements of F.

Proof Clearly $f(0) = 0$. Let $a \in K^*$. Since K^* is a group of order $q - 1$, we have $\mathrm{ord}_{K^*}(a) \mid (q-1)$ by Lagrange's theorem. In particular, $a^{q-1} = 1$, i.e., $f(a) = a^q - a = 0$. Therefore, each of the q elements of K is a root of f and consequently K is the splitting field of f. The last assertion in the statement of the theorem follows from the uniqueness of splitting fields (Proposition 1.71). ◀

This uniqueness allows us to talk about *the* finite field of cardinality q (rather than *a* finite field of cardinality q). We denote this (unique) field by $\mathbb{F}_q$.

Theorem 1.74 can be readily generalized for arbitrary extensions $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$, where $q = p^n, p \in \mathbb{P}, n, m \in \mathbb{N}$ (Exercise 1.5.2).

1.75 Proposition Let $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$, $m \in \mathbb{N}$, be a (finite) extension. There is a unique intermediate field with q^d elements, $d \in \mathbb{N}$, if and only if $d \mid m$. Furthermore, if $d \mid m$, then $\alpha \in \mathbb{F}_{q^m}$ belongs to the (unique intermediate) field $\mathbb{F}_{q^d}$, if and only if $\alpha^{q^d} = \alpha$.

Proof For any (positive) divisor d of m the splitting field L of $X^{q^d} - X$ consists of q^d elements and satisfies $\mathbb{F}_q \subseteq L \subseteq \mathbb{F}_{q^m}$. If $L' \neq L$ is another intermediate field with q^d elements, then there are more than q^d elements of $\mathbb{F}_{q^m}$, that are roots of $X^{q^d} - X$, a contradiction. Conversely, if L is an intermediate field, then L contains q^d elements, where $d = [L : \mathbb{F}_q]$. Since $m = [\mathbb{F}_{q^m} : \mathbb{F}_q] = [\mathbb{F}_{q^m} : L][L : \mathbb{F}_q]$, $d \mid m$. ◀

1.76 Corollary Let $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$, $m \in \mathbb{N}$, be a (finite) extension of finite fields, $\alpha \in \mathbb{F}_{q^m}$ and let $f(X) := \mathrm{minpoly}_{\alpha,\mathbb{F}_q}(X) \in \mathbb{F}_q[X]$. Then $\deg f$ divides m.

Proof Consider the intermediate field $\mathbb{F}_q(\alpha) \cong \mathbb{F}_q[X]/\langle f \rangle \cong \mathbb{F}_{q^d}$, where $d := \deg f$. ◀

Now we are in a position to prove a very important fact about the multiplicative group of a finite field.

1.77 Theorem Let K be a field (not necessarily finite). Then any *finite* subgroup G of the multiplicative group K^* is cyclic. In particular, $\mathbb{F}_q^*$ is cyclic.

Proof Since K is a field, for any $n \in \mathbb{N}$ the polynomial $X^n - 1$ has at most n roots in K and hence in G. The theorem then follows immediately from Exercise 1.5.4. ◀

1.78 Corollary Every finite extension $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$ of finite fields is simple.

Proof Let α be a generator of the cyclic group $\mathbb{F}_{q^m}^*$. Then m is the smallest of the positive integers s for which $\alpha^{q^s} = \alpha$. If f is the minimal polynomial of α over $\mathbb{F}_q$, then $\mathbb{F}_q \subseteq \mathbb{F}_q(\alpha) \cong \mathbb{F}_q[X]/\langle f \rangle \cong \mathbb{F}_{q^d} \subseteq \mathbb{F}_{q^m}$, where $d := \deg f$. Since $\alpha \in \mathbb{F}_{q^d}$, $\alpha^{q^d} = \alpha$, and hence we must have $d = m$, i.e., $\mathbb{F}_{q^m} = \mathbb{F}_q(\alpha)$. ◀

1.79 Corollary For any finite field $\mathbb{F}_q$ and $m \in \mathbb{N}$ there exists an irreducible polynomial $f \in \mathbb{F}_q[X]$ with $\deg f = m$.

Proof The minimal polynomial over $\mathbb{F}_q$ of a generator of $\mathbb{F}_{q^m}^*$ is irreducible in $\mathbb{F}_q[X]$ and has degree m. ◀

We now study some interesting properties of polynomials over finite fields. As before we concentrate on the polynomials in $\mathbb{F}_q[X]$ for an arbitrary $q = p^n$, $p \in \mathbb{P}$, $n \in \mathbb{N}$. We have seen how the polynomials $X^{q^m} - X$ proved to be important for understanding the structures of finite fields. But that's not all; these polynomials indeed have further roles to play. Therefore, we reserve the following special symbol: $\mathcal{I}_{q,m}(X) := X^{q^m} - X \in \mathbb{F}_q[X]$.

Let $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$ be a finite extension of finite fields and let $\alpha \in \mathbb{F}_{q^m}$ be a root of the polynomial $f(X) := a_n X^n + a_{n-1}X^{n-1} + \cdots + a_0 \in \mathbb{F}_q[X]$. Since each $a_i \in \mathbb{F}_q$, it follows that $a_i^q = a_i$. Therefore, $f(\alpha^q) = a_n\alpha^{qn} + a_{n-1}\alpha^{q(n-1)} + \cdots + a_0 = a_n^q\alpha^{qn} + a_{n-1}^q\alpha^{q(n-1)} + \cdots + a_0^q = f(\alpha)^q = 0$. More generally, for every $r = 0, 1, 2, \ldots$ the element α^{q^r} is a root of $f(X)$. We will now show that if f is irreducible in $\mathbb{F}_q[X]$, then all the roots of f are of this form. First let us prove the following important lemma:

1.80 Lemma Let $f(X) \in \mathbb{F}_q[X]$ be a non-constant irreducible polynomial. If f has a root in $\mathbb{F}_{q^m}$, then all the roots of f are in $\mathbb{F}_{q^m}$.

Proof Let $\alpha \in \mathbb{F}_{q^m}$ be a root of f. Then $f(X) = \mathrm{minpoly}_{\alpha,\mathbb{F}_q}(X)$. Since $\alpha^{q^m} = \alpha$, $f(X) \mid (X^{q^m} - X)$, i.e., any root β of f also satisfies $\beta^{q^m} = \beta$, i.e., $\beta \in \mathbb{F}_{q^m}$. ◀

1.81 Corollary The minimal polynomial of $\alpha \in \mathbb{F}_{q^m}$ over $\mathbb{F}_q$ is $(X - \alpha)(X - \alpha^q)\cdots(X - \alpha^{q^{d-1}})$, where d is the smallest of the integers $s \in \mathbb{N}$ for which $\alpha^{q^s} = \alpha$.

Proof Let $f(X) := \mathrm{minpoly}_{\alpha,\mathbb{F}_q} \in \mathbb{F}_q[X]$ and let $\delta := \deg f$. Then $\mathbb{F}_q(\alpha) \cong \mathbb{F}_q[X]/\langle f\rangle \cong \mathbb{F}_{q^\delta}$ is the smallest field containing ($\mathbb{F}_q$ and) α and hence all the roots of f. It follows that $\alpha^{q^s} = \alpha$ for $s = \delta$ and for no smaller positive integer values of s. Therefore, $\delta = d$ and all the conjugates of α are precisely $\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$. (One can easily check that $\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$ are all distinct.) ◀

1.82 Theorem The polynomial $\mathcal{I}_{q,m}(X) = X^{q^m} - X$ is the product of all non-constant monic irreducible polynomials in $\mathbb{F}_q[X]$ whose degrees divide m.

Proof We have the factorization $\mathcal{I}_{q,m}(X) = \prod_{\alpha\in\mathbb{F}_{q^m}}(X - \alpha)$ over $\mathbb{F}_{q^m}$. Now by Corollary 1.81 the minimal polynomial $f_\alpha(X)$ for every $\alpha \in \mathbb{F}_{q^m}$ over $\mathbb{F}_q$ divides $\mathcal{I}_{q,m}(X)$. By Corollary 1.76 we have $\deg(f_\alpha) \mid m$. Finally since $f_\alpha(X) = f_\beta(X)$ or $\gcd(f_\alpha(X), f_\beta(X)) = 1$ depending on whether α and β are conjugates or not, it follows that $\mathcal{I}_{q,m}(X)$ is *a* product of monic irreducible polynomials of $\mathbb{F}_q[X]$ whose degrees divide m. In order to show that $\mathcal{I}_{q,m}(X)$ is *the* product of all such polynomials, let us consider an arbitrary polynomial $g(X) \in \mathbb{F}_q[X]$ which is monic, irreducible over $\mathbb{F}_q$ and has degree $d \mid m$. Finite fields being perfect (Exercise 1.5.6), g has no multiple roots. Moreover, g has one (and hence all) roots in $\mathbb{F}_{q^d} \cong \mathbb{F}_q[X]/\langle g(X)\rangle$. Since $d \mid m$, we conclude from Proposition 1.75 that $\mathbb{F}_{q^d}$ is contained in $\mathbb{F}_{q^m}$. Thus g splits over $\mathbb{F}_{q^m}$ as well and, in particular, divides $\mathcal{I}_{q,m}$. ◀

An important consequence of the last theorem is that it leads to a procedure for checking the irreducibility of a polynomial $f(X) \in \mathbb{F}_q[X]$. Let $d := \deg f$. If $f(X)$ is reducible, it admits an irreducible factor of degree $\leqslant \lfloor d/2\rfloor$. Now $g_m := \gcd(f, \mathcal{I}_{q,m})$ is the product of all distinct irreducible factors of f whose degrees divide m. If all the gcds $g_1, \ldots, g_{\lfloor d/2\rfloor}$ are 1, f is irreducible. Otherwise f is reducible.

We end this section by explaining how elements of a finite field can be represented. Since $\mathbb{F}_{q^m}$ is a vector space of dimension m over $\mathbb{F}_q$, we can choose $\beta_0, \ldots, \beta_{m-1} \in \mathbb{F}_{q^m}$ that form an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^m}$. Each element $a \in \mathbb{F}_{q^m}$ then has a unique representation $a = a_0\beta_0 + \cdots + a_{m-1}\beta_{m-1}$, where each $a_i \in \mathbb{F}_q$. Therefore, if we have a representation for the elements of $\mathbb{F}_q$, we have the same for the elements of $\mathbb{F}_{q^m}$.

It is, then, easy to see that elements of any finite field can be represented, if we have representations of elements of prime fields. But we have the standard representation of $\mathbb{F}_p$ as the set $\{0, 1, \ldots, p-1\}$ with arithmetic modulo p.

So our problem now reduces to selecting a suitable basis $\beta_0, \ldots, \beta_{m-1}$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. To see how we can do that, let's choose *a priori* a fixed monic irreducible polynomial $f(X) \in \mathbb{F}_q[X]$ with $\deg f = m$. We represent $\mathbb{F}_{q^m}$ as $\mathbb{F}_q[X]/\langle f\rangle \cong \mathbb{F}_q(\alpha)$, where α (the residue class of X) is a root of f in $\mathbb{F}_{q^m}$. The elements $1, \alpha, \ldots, \alpha^{m-1} \in \mathbb{F}_{q^m}$ are linearly independent over $\mathbb{F}_q$, since otherwise we would have a polynomial of degree less than m of which α is a root. Therefore, $1, \alpha, \ldots, \alpha^{m-1}$ is an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^m}$, called the polynomial basis (with respect to the defining polynomial f). Elements of $\mathbb{F}_{q^m}$ are then polynomials in $\mathbb{F}_q[X]$ of degrees $< m$. Arithmetic in $\mathbb{F}_{q^m}$ is carried out as the polynomial arithmetic of $\mathbb{F}_q[X]$ modulo the irreducible polynomial f.

1.83 Example The elements of $\mathbb{F}_2$ are 0 and 1 with $0+0=0, 0+1=1, 1+0=1, 1+1=0$, $0\times 0 = 1\times 0 = 0\times 1 = 0$ and $1\times 1 = 1$. In order to represent $\mathbb{F}_8 = \mathbb{F}_{2^3}$ we choose the irreducible polynomial $f(X) = X^3 + X^2 + 1 \in \mathbb{F}_2[X]$. The elements of $\mathbb{F}_8$ are $a_2\alpha^2 + a_1\alpha + a_0$, where $a_i \in \{0, 1\}$. In order to demonstrate the arithmetic in $\mathbb{F}_8$ we take $a := \alpha^2 + 1, b := \alpha^2 + \alpha \in \mathbb{F}_8$. Their sum in $\mathbb{F}_8$ is $a + b = \alpha + 1$. On the other hand, the product is $ab = \alpha^4 + \alpha^3 + \alpha^2 + \alpha = \alpha(\alpha^3 + \alpha^2 + 1) + \alpha^2 = \alpha \cdot 0 + \alpha^2 = \alpha^2$.

Polynomial bases are the ones most commonly used in finite field implementations. However, there are other types of bases that are sometimes used.

Exercises for Section 1.5

1. Let F be a field (not necessarily finite) of characteristic $p \in \mathbb{P}$. Show that for every $a, b \in F$ we have $(a+b)^p = a^p + b^p$. (**Hint:** Use the binomial theorem.) More generally, prove that for $n \in \mathbb{N}$ and $a, b \in F$ we have $(a+b)^{p^n} = a^{p^n} + b^{p^n}$. (**Hint:** Use induction on n.)

2. Let p be a prime, $n, m \in \mathbb{N}$ and $q := p^n$. Let $F \subseteq K$ be an extension of finite fields with $|F| = q$ and $|K| = q^m$. Show that K is the splitting field of the polynomial $X^{q^m} - X \in F[X]$ over F. (**Hint:** Follow the proof of Theorem 1.74.)

3. [Solving this exercise requires the knowledge of Sylow subgroups and internal direct products of groups. If the reader is not already familiar with these topics, (s)he may consult any text-book on groups and/or algebra.]

Let G be a finite (multiplicatively written) Abelian group with identity e and of order $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where p_i are distinct primes and $\alpha_i \in \mathbb{N}$. For each i let H_i be the p_i-Sylow subgroup of G. Show that:

(a) $G = H_1 \cdots H_r$. (**Hint:** Let $i \neq j$ and $g \in H_i \cap H_j$. Then $\operatorname{ord} g$ divides both $p_i^{\alpha_i}$ and $p_j^{\alpha_j}$ and so is equal to 1, that is, $g = e$. Now let $h_i, h_i' \in H_i$ and $h_j, h_j' \in H_j$ with $h_i h_j = h_i' h_j'$. But then $h_i^{-1} h_i' = h_j (h_j')^{-1} \in H_i \cap H_j = \{e\}$. Thus $|H_i H_j| = |H_i||H_j|$. Generalize this argument to show that $|H_1 \cdots H_r| = n$.)

(b) Every element $g \in G$ can be written *uniquely* as $g = h_1 \cdots h_r$ with $h_i \in H_i$. Moreover, in that case we have $\operatorname{ord}_G g = (\operatorname{ord}_{H_1} h_1) \cdots (\operatorname{ord}_{H_r} h_r)$.

(c) G is cyclic, if and only if all of $H_1, \ldots, H_r$ are cyclic.

4. Let G be a finite (multiplicatively written) Abelian group with identity e. Assume that for every $n \in \mathbb{N}$ there are at most n elements x of G satisfying $x^n = e$. Show that G is cyclic. (**Hint:** First consider the special case $|G| = p^r$ for $p \in \mathbb{P}$ and $r \in \mathbb{N}$. Then each element $g \in G$ has order of the form p^{s_g} for some $s_g \in \{0, 1, \ldots, r\}$. Let s be the maximum of the integers s_g, $g \in G$. Show that $s = r$. This proves the assertion for the special case. For the general case use this special case in conjunction with Exercise 1.5.3.)

5. Let $F := \mathbb{F}_q$, $q = p^n$, $p \in \mathbb{P}$, $n \in \mathbb{N}$. Show that every element $\alpha \in F$ has a p-th root in F. (**Hint:** $\alpha^{p^n} = \alpha$.)

6. A field F is called perfect, if every (non-constant) irreducible polynomial in $F[X]$ has no multiple roots (in any extension of F).

(a) Show that if $\mathrm{char}\, F = 0$, then F is perfect.

(b) Let $\mathrm{char}\, F = p > 0$. Show that F is perfect, if and only if every element of F has a p-th root in F. In particular, finite fields are perfect.

(**Hint:** Use Exercise 1.4.3(e).)

7. Let $f(X) \in \mathbb{F}_q[X]$ be irreducible with $d := \deg f > 0$. Consider the extension $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$ and let $r := \gcd(d, m)$. Show that f is irreducible over $\mathbb{F}_{q^m}$, if and only if $r = 1$. (**Hint:** Assume $r > 1$. We have the tower of extensions $\mathbb{F}_q \subseteq \mathbb{F}_{q^r} \subseteq \mathbb{F}_{q^d}$ and $\mathbb{F}_{q^d}$ is the splitting field of f over $\mathbb{F}_q$ and hence over $\mathbb{F}_{q^r}$. Consider the minimal polynomial of a root $\alpha \in \mathbb{F}_{q^d}$ of f over $\mathbb{F}_{q^r}$. Conversely, let f be reducible over $\mathbb{F}_{q^m}$. Choose a non-constant irreducible factor $h \in \mathbb{F}_{q^m}[X]$ of f with $s := \deg h < d$. Now f has one (and hence all) roots in $\mathbb{F}_{q^{sm}}$ and, therefore, $d \mid sm$.)

8. Show that $\alpha \in \mathbb{F}_q^*$ is a primitive element (i.e., a generator) of $\mathbb{F}_q^*$, if and only if $\alpha^{(q-1)/\pi} \neq 1$ for every prime divisor π of $q - 1$. Find a primitive element of $\mathbb{F}_{16}^*$ with $\mathbb{F}_{16}$ represented as $\mathbb{F}_2[X]/\langle X^4 + X + 1\rangle$.