Algebraic Structures

Groningen, 2nd year bachelor mathematics, 2017
(mostly a translation from original Dutch lecture notes of L.N.M. van Geemen, H.W. Lenstra, F. Oort, and J. Top).

J. Top

Contents

preface

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

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

The translation project consists of two parts. The first one (*Algebraic Structures*) deals with the chapters $1-5$, $7-9$, and 12 of the Dutch notes. Many corrections and suggestions for improvement were offered by Dr. Max Kronberg who taught a course following these notes in the spring of 2017, and by Petra Hogeboom and Manoy Trip who at that time were students in this course. In the spring of 2019 student Wout Moltmaker mentioned a number of further small issues in the text, which have now been corrected. I am very grateful to all of them; needless to say that any mistakes and unclear parts in the exposition are only my fault. The second part of the translation project (which provides the material for the course *Advanced Algebraic Structures* and is not included here) discusses the chapters 8 (in slightly more detail than originally, to facilitate treating Galois theory), and chapters $6, 13$ on (projective) modules, then a discussion of basic Galois theory not present in the original notes, and finally the chapters 11 and 10 as well as an extended version of chapter 14 and a brief discussion of tensor products.

Groningen, January 2017 – April 2019
Jaap Top

I RINGS

I.1 Definition, examples, elementary properties

I.1.1 Definition. A *ring* (with 1) (also called unitary ring) is a five tuple $(R,+,\cdot,0,1)$ with R a set, $+$ and $\cdot$ maps written as:

$$+ : R\times R\to R,\quad (a,b)\mapsto a+b \qquad \cdot : R\times R\to R,\quad (a,b)\mapsto ab,$$

and 0 and 1 elements of R, such that the following properties (R1) to (R4) hold:

(R1) $(R,+,0)$ is an *abelian group*; this means:

- **(G1)** $a+(b+c)=(a+b)+c$ for all $a,b,c\in R$;
- **(G2)** $0+a=a+0=a$ for all $a\in R$;
- **(G3)** every $a\in R$ has an 'opposite' $-a\in R$ satisfying $a+(-a)=(-a)+a=0$;
- **(G4)** $a+b=b+a$ for all $a,b\in R$.

(R2) $a(bc)=(ab)c$ for all $a,b,c\in R$ (*associativity* of $\cdot$);

(R3) $a(b+c)=(ab)+(ac)$ and $(b+c)a=(ba)+(ca)$ for all $a,b,c\in R$ (the *distributive* laws).

(R4) $1a=a1=a$ for all $a\in R$.

A ring R is called *commutative* if moreover (R5) holds:

(R5) $ab=ba$ for all $a,b\in R$.

If $a,b\in R$ then $a+b$ and ab are called the *sum* and the *product* of a and b; the product ab is also denoted as $a\cdot b$. The maps $+$ and $\cdot$ are called the *addition* and the *multiplication* in R. If $(R,+,\cdot,0,1)$ is a ring then one says that R is a ring with addition $+$, multiplication $\cdot$, zero element 0, and unit element 1.

A trivial example of a ring is the *zero ring* $(\{0\},+,\cdot,0,0)$, with $0+0=0\cdot 0=0$. This is the only ring having $1=0$.

Some textbooks define 'rings' $(R,+,\cdot,0)$ only satisfying (R1), (R2), and (R3); such 'rings' are called non-unitary rings.

A *division ring* (or *skew field*) is a ring R such that in addition to (R1) to (R4), also (R6) holds:

(R6) $1\neq 0$, and for all $a\in R, a\neq 0$ there exists an inverse $a^{-1}\in R$ satisfying $a\cdot a^{-1}=a^{-1}\cdot a=1$.

A *field* (French: corps; German: Körper, Dutch: lichaam, Flemish: veld) is a *commutative* division ring (so (R1) to (R6) hold). A simple example of a field is the set $\{0,1\}$ with addition as in the abelian group $\mathbb{Z}/2\mathbb{Z}$ and product $0\cdot 0 = 0\cdot 1 = 1\cdot 0 = 0$ and $1\cdot 1 = 1$. The unit element is $1\,(\neq 0)$, this field we denote by $\mathbb{F}_2$.

I.1.2 Example. The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ of integers and rational, real, complex numbers (respectively) are with the familiar addition and multiplication rings. Moreover $\mathbb{Z},\mathbb{Q},\mathbb{R}$, and $\mathbb{C}$ are commutative. The rings $\mathbb{Q},\mathbb{R}$, and $\mathbb{C}$ are fields, and $\mathbb{Z}$ is not a field (condition (R6) is not satisfied in $\mathbb{Z}$). ∎

I.1.3 Example. Fix $n \in \mathbb{Z}_{>0}$. The set $\mathbb{Z}/n\mathbb{Z} = \{\overline{0},\overline{1},\ldots,\overline{n-1}\}$ with $\overline{i} = i + n\mathbb{Z} \subset \mathbb{Z}$, is equipped with an addition, since the elements $\overline{i}$ are the residue classes with respect to the normal subgroup $n\mathbb{Z} \subset \mathbb{Z}$. The rule

$$\overline{a}\cdot\overline{b} := \overline{a\cdot b},$$

with $a\cdot b$ the familiar multiplication in $\mathbb{Z}$ defines a product (verify for yourself that this is well-defined: if $\overline{a} = \overline{a_1}$ and $\overline{b} = \overline{b_1}$, then indeed $\overline{a\cdot b} = \overline{a_1\cdot b_1}$).

With respect to this addition and multiplication $\mathbb{Z}/n\mathbb{Z}$ is a commutative ring with unit element $\overline{1}$. In I.2.11 we will see that $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is a prime number. For $n = 1$ we have $\mathbb{Z}/1\mathbb{Z}$ which is the zero ring (hence, it is not a field). ∎

I.1.4 Example. Let $n \in \mathbb{Z}_{\geq 0}$. The set $\mathrm{M}(n,\mathbb{R})$ of all real $n \times n$-matrices is a unitary ring with the familiar matrix addition and matrix multiplication. For $n \geq 2$ this ring is not commutative.
In an analogous way one defines the ring $\mathrm{M}(n,R)$ for an arbitrary ring R and $n \in \mathbb{Z}_{\geq 1}$. ∎

I.1.5 Example. There are non-commutative division rings. Let K be a field (for example $\mathbb{R}$ or $\mathbb{Q}$) and take $\alpha,\beta \in K - \{0\}$. The *quaternion algebra* $(\alpha,\beta)_K$ consists of expressions (quaternions)

$$a + bi + cj + dk, \qquad \text{with} \quad a,b,c,d \in K.$$

We say two quaternions are equal only when their components are:

$$a + bi + cj + dk = a' + b'i + c'j + d'k \iff a = a',\ b = b',\ c = c',\ d = d'.$$

The quaternion algebra is a ring, as follows. Quaternion are added componentwise:

$$(a + bi + cj + dk) + (a' + b'i + c'j + d'k) = (a + a') + (b + b')i + (c + c')j + (d + d')k.$$

The multiplication of quaternions is based on the rules

$$ij = -ji = k, \quad i^2 = \alpha, \quad j^2 = \beta, \qquad \text{and}$$

$$x(a + bi + cj + dk) = (a + bi + cj + dk)x = ax + bxi + cxj + dxk$$

for $x = x + 0\cdot i + 0\cdot j + 0\cdot k \in K$. To obtain a ring one certainly needs

$$\begin{gathered} k^2 = (ij)(ij) = ((ij)i)j = (i(ji))j = -i^2j^2 = -\alpha\beta, \\ ik = i(ij) = \alpha j, \\ ki = (-ji)i = -\alpha j, \\ jk = j(-ji) = -\beta i, \\ kj = (ij)j = \beta i. \end{gathered}$$

These equalities lead to

$$\begin{aligned}&(a+bi+cj+dk)\cdot(a'+b'i+c'j+d'k) \quad = \\ &=(aa'+bb'\alpha+cc'\beta-dd'\alpha\beta)\\ &+(ab'+ba'-cd'\beta+dc'\beta)i\\ &+(ac'+bd'\alpha+ca'-db'\alpha)j\\ &+(ad'+bc'-cb'+da')k.\end{aligned}$$

A straightforward verification shows that in this way indeed a ring is defined.

Given $q=a+bi+cj+dk\in(\alpha,\beta)_K$, put

$$\overline{q}:=a-bi-cj-dk.$$

We define

$$N(q):=q\overline{q}=(a+bi+cj+dk)(a-bi-cj-dk)=a^2-\alpha b^2-\beta c^2+\alpha\beta d^2.$$

In particular we have $N(q)\in K$ for all quaternions q. Observe that

$$N(q)\neq 0\Rightarrow(a+bi+cj+dk)^{-1}=\frac{1}{N(q)}\bar{q}=\frac{a}{N(q)}-\frac{b}{N(q)}i-\frac{c}{N(q)}j-\frac{d}{N(q)}k.$$

As a consequence

$$(\alpha,\beta)_K \quad \text{is a } \textit{division ring} \text{ if and only if}$$

for all $a,b,c,d\in K$ it holds that

$$N(a+bi+cj+dk)=0\implies a=b=c=d=0.$$

Namely, if $N(q)=0\Rightarrow q=0$, then $q\neq 0$ has as inverse $q^{-1}:=\frac{1}{N(q)}\overline{q}$ and therefore $(\alpha,\beta)_K$ is a division ring. Vice versa, if $q\neq 0$ exists with $N(q)=0$, then $q\overline{q}=0$. Should an inverse q^{-1} of q exist, then $0=q^{-1}q\overline{q}=1\cdot\overline{q}=\overline{q}$ which implies $q=0$, a contradiction. Hence our q has no inverse which shows $(\alpha,\beta)_K$ is not a division ring.

In the special case $K=\mathbb{R}$ quaternions were introduced in 1843 by Hamilton (Sir William Rowan Hamilton, English-Irish mathematician, 1805-1865) as the quaternion algebra $(-1,-1)_{\mathbb{R}}$; one writes

$$\mathbb{H}:==(-1,-1)_{\mathbb{R}},\qquad \text{so in } \mathbb{H}:\quad i^2=j^2=k^2=-1.$$

In particular $\mathbb{H}$ is a (non-commutative) division ring, since $q=a+bi+cj+dk$ satisfies $N(q)=a^2+b^2+c^2+d^2=0$ for $a,b,c,d\in\mathbb{R}$ precisely when $a=b=c=d=0$.

We now present a quaternion algebra which is not a division ring. In $\mathrm{M}(2,\mathbb{R})$ take the matrices

$$i:=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\qquad j:=\begin{pmatrix}0&1\\1&0\end{pmatrix},\qquad k:=ij=-ji:=\begin{pmatrix}0&1\\-1&0\end{pmatrix}.$$

Note that in $\mathrm{M}(2,\mathbb{R})$ we have

$$\begin{pmatrix}p&q\\r&s\end{pmatrix}=\frac{p+s}{2}\begin{pmatrix}1&0\\0&1\end{pmatrix}+\frac{p-s}{2}\begin{pmatrix}1&0\\0&-1\end{pmatrix}+\frac{q+r}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix}+\frac{q-r}{2}\begin{pmatrix}0&1\\-1&0\end{pmatrix}.$$

In this way $\mathrm{M}(2,\mathbb{R})$ is identified with the quaternion algebra $(1,1)_{\mathbb{R}}$. This is not a division ring: for instance $j+k\neq 0$ but $N(j+k)=0$, so $(j+k)$ has no inverse in the quaternion algebra $(1,1)_{\mathbb{R}}$. ∎

We refer to Section I.3 for more examples of rings and for other methods to construct rings.

I.1.6 Definition. A subset R' of a ring R is called a (unitary) *subring* of R if (D1), (D2), and (D3) hold:

(D1) $1\in R'$;
(D2) R' is a subgroup of the additive group of R, i.e., $a+(-b)\in R'$ for all $a,b\in R'$;
(D3) $ab\in R'$ for all $a,b\in R'$.

A subring R' of a ring R is itself a ring, with the addition and multiplication of R. If R is commutative, then so is R'.

A trivial example of a subring of R is R itself.

I.1.7 Example. Let $i \in \mathbb{C}$ satisfy $i^2 = -1$. The set $\mathbb{Z}[i] = \{a+bi : a,b \in \mathbb{Z}\} \subset \mathbb{C}$ is with the usual addition and multiplication of complex numbers a ring, hence a subring of $\mathbb{C}$. We call $\mathbb{Z}[i]$ the *ring of Gaussian integers*. It is a commutative ring with 1, but not a field. The set $\mathbb{Q}[i] = \{a+bi : a,b \in \mathbb{Q}\}$ is also a subring of $\mathbb{C}$ and it is even a field: the inverse of $a+bi$ ($\neq 0$) is given by $\frac{a}{a^2+b^2} + \frac{-b}{a^2+b^2}\,i \in \mathbb{Q}[i]$. Analogous remarks apply to

$$\mathbb{Z}[\sqrt{m}] = \{a+b\sqrt{m} : a,b \in \mathbb{Z}\},$$
$$\mathbb{Q}[\sqrt{m}] = \{a+b\sqrt{m} : a,b \in \mathbb{Q}\},$$

with m an integer that is not the square of some integer (in the special case $m = -1$ one recovers $\mathbb{Z}[i]$ and $\mathbb{Q}[i]$). ▬

I.1.8 Notation. If R is a ring and $a,b \in R$, then one abbreviates the sum $a+(-b)$ as $a-b$. So this is the sum of a and the opposite (additive inverse) of b.

I.1.9 Theorem. *Let R be a ring. For $a,b,b_1,\ldots,b_n,c \in R$ we have*

$$\begin{aligned} a(b_1+b_2+\ldots+b_n) &= ab_1+ab_2+\ldots+ab_n, \\ (b_1+b_2+\ldots+b_n)a &= b_1a+b_2a+\ldots+b_na, \\ a(b-c) &= ab-ac \\ a\cdot 0 &= 0\cdot a = 0. \end{aligned}$$

Proof. The first two equalities follow from the distributive law (R3) using mathematical induction w.r.t. n. Since

$$a(b-c)+ac = a((b-c)+c) = a(b+(-c)+c) = ab$$

one concludes $a(b-c) = ab-ac$. Finally

$$a\cdot 0 = a\cdot(0-0) = a\cdot 0 - a\cdot 0 = 0$$

and analogously $0\cdot a = 0$. This proves the theorem. ■

I.2 Units and zero divisors

By (R1) every ring R is an abelian group w.r.t. addition. This group is sometimes denoted by R^+; so R^+ is the same set R, with the same addition as in R, however one 'forgets' the multiplication.
With respect to the multiplication the unitary ring R only yields a group $(R,\cdot,1)$ in case $R = \{0\}$. In spite of this, the next definition allows one to talk about a multiplicative group of a ring.

I.2.1 Definition. Let R be a ring with 1. An element $a \in R$ is called a *unit* (or, invertible) if some $b \in R$ exists such that

$$ab = ba = 1.$$

(Observe the somewhat peculiar terminology: the unit element is indeed a unit, but vice versa a unit need not be equal to the unit element.) The set of units in R is denoted $R^\times$ and is called the *unit group* of R (it is indeed a group as will be shown in Theorem I.2.3). Some texts also use the notation $U(R)$.

An element $a \in R$ is called a *left unit* if $\quad \exists b \in R :\ ab = 1,$
and it is called a *right unit* if $\quad \exists c \in R :\ ca = 1.$

If $a \in R$ is both a left and a right unit, then a is a unit: namely,

$$ab = 1, \quad ca = 1 \quad \Longrightarrow \quad cab = c \Longrightarrow b = c.$$

In a commutative ring the notion left unit (or right unit) of course coincides with 'unit', but in a non-commutative ring a left unit is not necessarily also a right unit: see I.3.

I.2.2 Example. $\mathbb{Z}^\times = \{1,-1\}$, $\mathbb{Q}^\times = \mathbb{Q} - \{0\}$, $\mathbb{R}^\times = \mathbb{R} - \{0\}$, $\mathbb{C}^\times = \mathbb{C} - \{0\}$, $\mathbb{H}^\times = \mathbb{H} - \{0\}$. ▁■

In general, see (R6): R is a division ring $\iff$ $R^\times = R - \{0\}$.

I.2.3 Theorem. *The unit group $R^\times$ of a ring R with 1 is a group with respect to multiplication.*

Proof. First we show that $ab \in R^\times$ in case $a, b \in R^\times$. Namely, if $a, b \in R^\times$ then $c, d \in R$ exist with $ac = ca = 1$ and $bd = db = 1$, hence $(ab)\cdot(dc) = (dc)\cdot(ab) = 1$ with $dc \in R$; so $ab \in R^\times$.

The associativity of the product follows immediately from (R2).

The set $R^\times$ has a neutral element, namely $1 \in R^\times$ satisfies $1 \cdot 1 = 1$, and (R4) shows that 1 also satisfies $a \cdot 1 = 1 \cdot a = a$.

Finally, if $a \in R^\times$ then $b \in R$ exists with $ab = ba = 1$; this b satisfies $b \in R^\times$ hence every element in $R^\times$ has an inverse in $R^\times$.

We have now verified the 4 axioms of a group, hence the theorem is proven. ■

In case R is commutative, obviously $R^\times$ is abelian. The converse is not true in general: one can construct non-commutative (unitary) rings R with $R^\times$ abelian, see for example exercise 18.

I.2.4 Example. If $A \in \mathrm{M}(n, \mathbb{R})$ is invertible with inverse B then $AB = BA = I$, where I denotes the identity matrix. Furthermore

A is a left unit $\iff$ A is a right unit $\iff$ $\det(A) \neq 0$.

So $\mathrm{M}(n,\mathbb{R})^\times = \mathrm{GL}(n,\mathbb{R})$ (in fact this is the definition of the group $\mathrm{GL}(n,\mathbb{R})$). Here we may replace $\mathbb{R}$ by an arbitrary (unitary) commutative ring. ▁■

I.2.5 Example. Take $R = \mathbb{Z}[\sqrt{m}]$ as in I.1.7, with m an integer that is not a square. We define the *norm*

$$N : R \longrightarrow \mathbb{Z}, \quad N(a + b\sqrt{m}) = (a + b\sqrt{m})\cdot(a - b\sqrt{m}) = a^2 - mb^2.$$

One easily verifies that $N(\alpha\beta) = N(\alpha)\cdot N(\beta)$ for all $\alpha, \beta \in R$, and $N(0) = 0$, $N(1) = 1$. We claim:

$$\alpha \in R^\times \iff N(\alpha) = \pm 1;$$

$\Leftarrow$: for $\alpha = a + b\sqrt{m}$ with $N(\alpha) = \pm 1$ we have $(a + b\sqrt{m})\cdot(a - b\sqrt{m}) = \pm 1$, hence $\pm(a - b\sqrt{m})$ is an inverse of α.

$\Rightarrow$: is $\alpha\beta = 1$ then $N(\alpha)\cdot N(\beta) = N(\alpha\beta) = N(1) = 1$, and $N(\alpha)$, $N(\beta) \in \mathbb{Z}$, hence $N(\alpha) = N(\beta) = \pm 1$.

This shows that the search for units in $\mathbb{Z}[\sqrt{m}]$ is equivalent to finding solutions of the equation

$$a^2 - m \cdot b^2 = \pm 1$$

in integers a, b.

For $m < 0$ solving this equation is not hard: it holds that $a^2 - m \cdot b^2 = a^2 + |m| \cdot b^2$, and because squares of integers are nonnegative, the latter expression is only equal to ± 1 in the cases

$$a = \pm 1,\ b = 0, \qquad \text{and}$$

$$a = 0,\ b = \pm 1,\ m = -1.$$

Therefore

$$\mathbb{Z}[i]^\times = \{1, i, -1, -i\} \qquad (\text{the case } m = -1),$$

$$\mathbb{Z}[\sqrt{m}]^\times = \{1, -1\} \qquad \text{if} \quad m < -1.$$

In case $m > 0$ (and not a square) the equation $x^2 - my^2 = \pm 1$ is much more interesting. One can show that for every such $m > 0$ the "Pell equation" $x^2 - my^2 = 1$ has a solution $x, y \in \mathbb{Z}_{>0}$. This yields a unit $\epsilon = x + y\sqrt{m} > 1$ of $R = \mathbb{Z}[\sqrt{m}]$, and infinitely many units in R are than obtained as $\ldots, \pm\epsilon^{-2}, \pm\epsilon^{-1}, \pm 1, \pm\epsilon, \pm\epsilon^2, \ldots$. Apparently Pell's equation has infinitely many solutions.
Example: If $m = 2$ then $x_1 = y_1 = 1$ is a solution of $x^2 - 2y^2 = \pm 1$, hence $\epsilon = 1 + \sqrt{2}$ is an element of $\mathbb{Z}[\sqrt{2}]^\times$. Considering $\epsilon^n, n \geq 0$, yields the solutions

$x_0 = 1$	$y_0 = 0$	$x_5 = 41$	$y_5 = 29$
$x_1 = 1$	$y_1 = 1$	$x_6 = 99$	$y_6 = 70$
$x_2 = 3$	$y_2 = 2$	$x_7 = 239$	$y_7 = 169$
$x_3 = 7$	$y_3 = 5$		
$x_4 = 17$	$y_4 = 12$		

(And more generally: $x_{n+1} = 2x_n + x_{n-1},\ y_{n+1} = 2y_n + y_{n-1}$.)

For $m = 67$ the 'simplest nontrivial unit' is the one with $x = 48842, y = 5967$. More information may be found in H. Davenport, The higher arithmetic, Ch.IV, section 11. This includes an explanation why the name of John Pell (1611-1685) is erroneously attached to the equation. ∎

In an arbitrary ring it could happen that $a \cdot b = 0$ while $a \neq 0$, $b \neq 0$. For example $\overline{2} \cdot \overline{3} = \overline{0}$ in $\mathbb{Z}/6\mathbb{Z}$. In $\mathbb{Z}/8\mathbb{Z}$ one even has $\overline{2}^3 = \overline{0}$.

I.2.6 Definition. An element a in a ring R is called a *left zero divisor* if:
$a \neq 0$ and $\exists b \in R : b \neq 0 \wedge ab = 0$;
the element a is called a *right zero divisor* if $a \neq 0$ and $\exists c \in R : c \neq 0 \wedge ca = 0$;
it is called a *zero divisor* if it is either a left- or a right zero divisor (or both).

A *nilpotent element* is an $a \in R$, $a \neq 0$, such that $a^n = 0$ for some $n \in \mathbb{N}$. In particular a nilpotent element is a zero divisor, both left and right.

An element $a \in R$ is called an *idempotent element* if $a^2 = a$ and $0 \neq a \neq 1$. An idempotent element is in particular a zero divisor (both left and right), because $a^2 = a$ implies $a(a-1) = (a-1)a = 0$ and $0 \neq a \neq 1$ implies $a, a - 1 \neq 0$.

I.2.7 Example. In $\mathrm{M}(2, \mathbb{R})$ consider

$$a := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad b := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad c := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad .$$

Check for yourself that $ab = 0$, hence a is a left zero divisor and b is a right zero divisor. Note that $ba \neq 0$, however $ca = 0$, so a is (anyway) a right zero divisor. Moreover $a^2 = 0$, hence a is a nilpotent element (this shows again that a is both a left- and a right zero divisor).

Note also that $b^2 = b$ and $c^2 = c$, so b and c are idempotent elements. ∎

I.2.8 Theorem. *An element a in a commutative ring R with 1 can not be both a zero divisor and a unit.*

Proof. Note that since R is assumed to be commutative, the notions of left- and right zero divisor coincide. Suppose a is a zero divisor: $a \neq 0$, and $ab = 0$ for some $b \in R, b \neq 0$; and moreover a is a unit: $ac = ca = 1$ ($c \in R$). Then $c \cdot a \cdot b = 1 \cdot b = b$ and also $c \cdot a \cdot b = c \cdot 0 = 0$, hence $b = 0$, a contradiction. This proves I.2.8. ∎

I.2.9 Remark. The proof of I.2.8 in fact shows that in an arbitrary (not necessarily commutative) ring a left zero divisor is not a right unit (see Definition I.2.1). In the same way a right zero divisor is not a left unit. In I.3 we will discuss an example of a left unit which is also a left zero divisor.

I.2.10 Corollary. *A division ring contains no zero divisors.*

Proof. This is a consequence of I.2.8, since all elements $\neq 0$ in a division ring are units. ∎

I.2.11 Theorem. *For $n \in \mathbb{Z}_{>0}$ one has: $\mathbb{Z}/n\mathbb{Z}$ is a field $\iff$ n is a prime number.*

Proof. For a commutative ring R with 1 it holds that R is a field $\Leftrightarrow R^\times = R - \{0\}$. If n is not prime, then either $n = 1$ and in that case $\overline{0} = \overline{1}$ in $\mathbb{Z}/1\mathbb{Z}$ which shows $\mathbb{Z}/n\mathbb{Z}$ is not a field, or $n = ab$ with $1 < a, b < n$. In the latter case

$$\bar{a}, \bar{b} \neq \bar{0} \in \mathbb{Z}/n\mathbb{Z} \qquad \text{and} \quad \bar{a}\bar{b} = \bar{n} = \bar{0}.$$

Hence by Theorem I.2.8 the element $\bar{a} \in \mathbb{Z}/n\mathbb{Z} - \{\bar{0}\}$ is not a unit, so $\mathbb{Z}/n\mathbb{Z}$ is not a field.

In case n is prime, suppose $\bar{a} \neq \bar{0}$ in $\mathbb{Z}/n\mathbb{Z}$. We have to show that $\bar{a}$ admits an inverse. Note that the additive group $(\mathbb{Z}/n\mathbb{Z})^+$ consists of n elements, a prime number. Hence the subgroup generated by $\bar{a}$ is all of $(\mathbb{Z}/n\mathbb{Z})^+$. Since $\bar{1} \in (\mathbb{Z}/n\mathbb{Z})^+$, an $m \in \mathbb{Z}_{>0}$ exists with $m\bar{a} := \bar{a} + \ldots + \bar{a}$ (m times) equal to $\bar{1}$. So $\bar{m}\bar{a} = \bar{a}\bar{m} = \bar{1}$, and $\bar{m}$ is the requested inverse of $\bar{a}$. This proves Theorem I.2.11. ∎

I.2.12 Notation. Given a prime number p, the field $\mathbb{Z}/p\mathbb{Z}$ is denoted by

$$\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}.$$

The field $\mathbb{F}_p$ consists of p (a prime number) elements. In Chapter IX we will construct for various other positive integers q a field $\mathbb{F}_q$ containing precisely q elements. If q is not prime, this field $\mathbb{F}_q$ is of course *not* equal to $\mathbb{Z}/q\mathbb{Z}$ because by Theorem I.2.11 the latter ring is not a field in case q is not a prime.

I.2.13 Definition. An *integral domain* (or *domain* or *integral ring*) is a commutative ring with $1 \neq 0$ having no zero divisors.

I.2.14 Example. Examples of domains are fields (by I.2.10), for instance

$$\mathbb{Q},\ \mathbb{R},\ \mathbb{C},\ \mathbb{F}_{59}.$$

Moreover (unitary) subrings of fields are domains, for instance $\mathbb{Z}$, $\mathbb{Z}[i]$. In I.3 we will see that every domain is a subring of a field.

Some rings that are *not* domains are $\mathbb{H}$ (not commutative), $\mathbb{Z}/1\mathbb{Z}$ ($1 = 0$), and $\mathbb{Z}/57\mathbb{Z}$ ($\overline{3} \cdot \overline{19} = \bar{0}$, so this ring contains zero divisors). ∎

I.2.15 Theorem. *Suppose R is a ring without zero divisors (for example, R could be a domain), and let $a, b, c \in R$. Then:*

(a) $ab = 0 \iff a = 0$ *or* $b = 0$,
(b) $ab = ac \iff a = 0$ *or* $b = c$.

Proof. (a) $\Leftarrow$ is a consequence of I.1.9; $\Rightarrow$: if $ab = 0$ and $a \neq 0 \neq b$, then a and b are zero divisors, a contradiction.

(b) We have $ab = ac \Leftrightarrow ab - ac = 0 \Leftrightarrow a(b-c) = 0$ (by I.1.9) $\Leftrightarrow a = 0$ or $b - c = 0$ (by (a) above) $\Leftrightarrow a = 0$ or $b = c$. This proves I.2.15.

■

I.3 Constructions of rings

We now present various important ways to construct rings.

Product of rings

If R_1 and R_2 are rings, then one defines a coordinatewise addition and multiplication on $R = R_1 \times R_2$ by

$$(r_1, r_2) + (s_1, s_2) = (r_1 + s_1, r_2 + s_2), \qquad (r_1, r_2) \cdot (s_1, s_2) = (r_1 s_1, r_2 s_2)$$

Here $r_1, s_1 \in R_1$, $r_2, s_2 \in R_2$. It is not hard to verify that this makes R a ring. Denoting the zero element and the unit element of R by 0_R and 1_R respectively, one concludes $0_R = (0,0)$ and $1_R = (1,1)$. The ring R is commutative if and only if both R_1 and R_2 are commutative. Moreover $R^\times = R_1^\times \times R_2^\times$. The proofs of these assertions are left as an exercise to the reader.

A ring $R_1 \times R_2$ with $R_1 \neq \{0\}$ and $R_2 \neq \{0\}$ necessarily contains zero divisors, since

$$(a,0) \cdot (0,b) = (a \cdot 0, 0 \cdot b) = (0,0),$$

for all $a \in R_1$, $b \in R_2$.

As a final remark, note that the elements $(1,0)$ and $(0,1)$ are idempotents in $R_1 \times R_2$.

Fields of fractions

Let R be a domain. We will construct a field called the *field of fractions* (or quotient field) of R, notation: $Q(R)$, with the following properties: first, R is contained in $Q(R)$. So in particular every $s \in R$ with $s \neq 0$ has an inverse $s^{-1} \in Q(R)$ (regardless of s being in $R^\times$ or not). Secondly, every element of $Q(R)$ can be written as $r \cdot s^{-1}$ for some $r, s \in R$, $s \neq 0$. The construction generalises the construction of $\mathbb{Q} = Q(\mathbb{Z})$ starting from $\mathbb{Z}$.

Let $S = R - \{0\}$. On the set $R \times S = \{(a,s) : a, s \in R, s \neq 0\}$ we define an equivalence relation $\sim$ by

$$(a,s) \sim (b,t) \quad \Longleftrightarrow \quad at = bs.$$

It is not hard to verify that indeed $\sim$ defines an equivalence relation: reflexivity ($(a,s) \sim (a,s)$) and symmetry ($(a,s) \sim (b,t) \Rightarrow (b,t) \sim (a,s)$) are trivial; transitivity ($(a,s) \sim (b,t) \wedge (b,t) \sim (c,u) \Rightarrow (a,s) \sim (c,u)$) is shown as follows.

From $(a,s) \sim (b,t)$ we deduce $at = bs$, hence $atu = bsu$. Now $(b,t) \sim (c,u)$ shows $bu = ct$, so $bus = cts$. Since R is commutative, $aut = atu = bsu = bus = cts = cst$. As

$$aut = cst \quad \Longrightarrow \quad (au - cs)t = 0,$$

and $t \neq 0$, it follows from I.2.13 (b) that $au = cs$. As a consequence $(a,s) \sim (c,u)$, which is what we wanted to show.

Now let $Q(R)$ denote the set of equivalence classes of $\sim$:

$$Q(R) = (R \times S)/\sim .$$

For the equivalence class containing (a,s) we introduce the suggestive notation $\frac{a}{s}$. So

$$Q(R) = \left\{ \frac{a}{s} : a, s \in R,\ s \neq 0 \right\}$$

and

$$\frac{a}{s} = \{(b,t) \in R \times S \,:\, (a,s) \sim (b,t)\}$$

and

$$\frac{a}{s} = \frac{b}{t} \quad \Longleftrightarrow \quad at = bs.$$

We define an addition and a multiplication on $Q(R)$ by

$$\frac{a}{s} + \frac{b}{t} = \frac{at+bs}{st} \qquad \text{(note : } st \neq 0 \text{ because } R \text{ is a domain),}$$
$$\frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st}.$$

Of course one needs to verify that these operations are well defined, in other words, they do not depend on the choice of a pair in the given equivalence class. In formulas: if $\frac{a'}{s'} = \frac{a}{s}$ and $\frac{b'}{t'} = \frac{b}{t}$ then we have to verify that $\frac{a't'+b's'}{s't'} = \frac{at+bs}{st}$ and $\frac{a'b'}{s't'} = \frac{ab}{st}$. This is indeed the case:

$$\frac{a'}{s'} = \frac{a}{s} \;\wedge\; \frac{b'}{t'} = \frac{b}{t} \quad \Longrightarrow \quad a's = as' \;\wedge\; b't = bt' \quad \Longrightarrow$$

$$\begin{aligned}(a't' + b's')st \quad &= a'st't + b'ts's = as't't + bt's's \\ &= (at+bs)s't' \quad \Longrightarrow \quad \tfrac{a't'+b's'}{s't'} = \tfrac{at+bs}{st}\end{aligned}$$

and for the product the verification is even less involved.

The verification that $Q(R)$ equipped with this addition and multiplication, and using the zero element $\frac{0}{1}$ and the unit element $\frac{1}{1}$ satisfies $(R1)$ up to $(R6)$ is somewhat time consuming but it yields no difficulties. We conclude that $Q(R)$ is a field.

We consider R as a subring of $Q(R)$ by identifying the element $a \in R$ with $\frac{a}{1} \in Q(R)$:

$$R \subset Q(R), \qquad r = \frac{r}{1}.$$

Note that in this way distinct elements of R remain distinct in $Q(R)$ since by definition $\frac{a}{1} = \frac{b}{1} \Leftrightarrow a \cdot 1 = b \cdot 1 \Leftrightarrow a = b$. Moreover $\frac{a}{1} + \frac{b}{1} = \frac{a \cdot 1 + b \cdot 1}{1 \cdot 1} = \frac{a+b}{1}$ and $\frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1}$, so indeed the elements $\{\frac{a}{1} \,:\, a \in R\}$ identified with R form a subring of the field $Q(R)$.

For $s \in S = R - \{0\}$ the inverse in $Q(R)$ is $\frac{1}{s}$: namely, in $Q(R)$ one has $\frac{1}{s} \cdot \frac{s}{1} = \frac{s}{s} = \frac{1}{1}$. So using slightly sloppy notation, one concludes $s^{-1} = \frac{1}{s}$. As a consequence, any $\frac{a}{s} \in Q(R)$ satisfies $\frac{a}{s} = \frac{a}{1} \cdot \frac{1}{s} = a \cdot s^{-1}$. This completes the construction of the field of fractions of R, satisfying the properties mentioned earlier. The assertion from Example I.2.14 stating that every domain is a subring of some field also follows from this construction, since $R \subset Q(R)$.

I.3.1 Example. Any field K gives rise to a ring of polynomials with coefficients in K in the variable X, called $K[X]$. The precise definition will be given in Chapter III.1. The ring $K[X]$ is a domain, and we define the *field of rational functions in one variable* over K by

$$K(X) := Q(K[X]).$$

Some elements of $K(X)$ are $\frac{1}{1+X} = \frac{X}{X+X^2}$ and $\frac{1-X^2}{1-X+X^3}$. ∎

For an important generalisation in the theory of commutative rings of the construction of fields of fractions we refer to Exercise 28.

Endomorphism rings.

Suppose $(A,+,0)$ is an abelian group, and let $\mathrm{End}(A)$ denote the set of endomorphisms of A:

$$\mathrm{End}(A) := \{f : A \to A : \ f(a+b) = f(a) + f(b) \quad \forall a,b \in A\}.$$

For $f,g \in \mathrm{End}(A)$ we define $f+g : A \to A$ and $fg : A \to A$ by

$$(f+g)(a) = f(a) + g(a), \quad fg(a) = f(g(a)).$$

The assumption that A is abelian assures that $f+g \in \mathrm{End}(A)$. We have $fg \in \mathrm{End}(A)$ as well. It is a straightforward verification that $\mathrm{End}(A)$ equipped with this addition and multiplication is a ring, called the *endomorphism ring* of A. It is a ring with unit element id_A, the identity map. The unit element is not equal to the zero element, except in the case $A = 0$.

I.3.2 Example. Let $A = \mathbb{R}^n$, with $n \in \mathbb{Z}_{>0}$. Considering any $n \times n$-matrix over $\mathbb{R}$ as an $\mathbb{R}$-linear endomorphism of the vector space A, the addition and multiplication of matrices correspond to the addition and multiplication of endomorphisms as defined above. In this way $\mathrm{M}(n,\mathbb{R})$ may be regarded as a subring of $\mathrm{End}(A)$.

As $\mathrm{M}(n,\mathbb{R})$ is non-commutative for $n \geq 2$, we conclude that $\mathrm{End}(A)$ is not commutative. So apparently abelian groups A exist for which $\mathrm{End}(A)$ is not a commutative ring. ∎

I.3.3 Example. Put $A = \mathbb{R}[X]^+$ (the additive group of the ring of polynomials over $\mathbb{R}$). Define $f, g, x \in \mathrm{End}(A)$ by

$$\begin{array}{lcl} f\colon a_0 + a_1X + \ldots + a_nX^n & \mapsto & a_1 + a_2X + \ldots + a_nX^{n-1}, \\ g\colon a_0 + a_1X + \ldots + a_nX^n & \mapsto & a_0, \\ x\colon a_0 + a_1X + \ldots + a_nX^n & \mapsto & a_0X + a_1X^2 \ldots + a_nX^{n+1}. \end{array}$$

Moreover write $1 = id_A$, the unit element of $\mathrm{End}(A)$. One computes without difficulty that in $\mathrm{End}(A)$ it holds that:

$$fx = 1, \quad fg = 0, \quad gx = 0.$$

So f is a left unit as well as a left zero divisor in $\mathrm{End}(A)$. By Remark I.2.9 it follows that f is neither a right unit nor a right zero divisor. Similarly x is a right unit and a right zero divisor, but not a left unit and not a left zero divisor.

Differentiation defines $d \in \mathrm{End}(A)$:

$$d : a_0 + a_1X + \ldots + a_nX^n \mapsto a_1 + 2a_2X + \ldots + na_nX^{n-1}.$$

Note that we have:

$$\begin{array}{ll} dx(a_0 + a_1X + \ldots + a_nX^n) & = d(a_0X + a_1X^2 + \ldots + a_nX^{n+1}) \\ & = a_0 + 2a_1X + \ldots + (n+1)a_nX^n, \end{array}$$

and

$$\begin{array}{ll} xd(a_0 + a_1X + \ldots + a_nX^n) & = x(a_1 + 2a_2X + \ldots + na_nX^{n-1}) \\ & = a_1X + 2a_2X^2 + \ldots + na_nX^n. \end{array}$$

Apparently any $f \in A$ satisfies $(dx - xd)f = f$, hence in the ring $\mathrm{End}(A)$ one finds:

$$dx - xd = id_A = 1.$$

We consider $\mathbb{R}$ as a subring of $\mathrm{End}(A)$ via

$$a : a_0 + a_1X + \ldots + a_nX^n \mapsto aa_0 + aa_1X + \ldots + aa_nX^n.$$

Since End(A) is a ring, every finite linear combination

$$\sum_{i,j}^{<\infty} a_{ij} x^i d^j, \qquad a_{ij} \in \mathbb{R}$$

is in End(A). Repeatedly applying the rule $dx - xd = 1$ one concludes that

$$W := \left\{ \sum_{i,j}^{<\infty} a_{ij} x^i d^j \in \mathrm{End}(A) \right\}$$

is a subring of End(A). One calls W the *Weyl algebra*; it consists of linear differential operators with real polynomials in x as coefficients. ∎

Rings of functions

Let V be a set and R a ring and $T = R^V$ the set of all maps from V to R. One makes T into a ring by using the pointwise sum and product of functions $f, g : V \to R$:

$$(f+g)(v) = f(v) + g(v) \in R,$$

$$(fg)(v) = f(v) \cdot g(v) \in R,$$

for any $v \in V$. In case $V = \{v_1, v_2, \ldots, v_n\}$ consisting of $n \in \mathbb{Z}_{>0}$ distinct elements, then R^V is 'the same' ring as $R \times R \times \ldots \times R$ (product of rings: see the beginning of I.3). Note that R^V contains zero divisors whenever $\#V \geq 2, R \neq \{0\}$: take $v \in V$ and $r \in R - \{0\}$ and define $f \in R^V$ by $f(v) = r$, $f(w) = 0$ for all $w \neq v$. Similarly take $g \in R^V$ defined by $g(v) = 0$ and $g(w) = r$ for all $w \neq v$. Then $g \neq 0 \neq f$ and $gf = fg = 0$.

Other interesting rings are obtained by putting additional conditions on the functions in T. As an example, take $V = [0,1]$ the closed interval between 0 and 1, and $R = \mathbb{R}$, and consider

$$C([0,1]) = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}.$$

This is a subring of the ring $\mathbb{R}^{[0,1]}$ defined above, and this subring also contains zero divisors: define $f, g \in C([0,1])$ by

$$f(x) = \begin{cases} x - \frac{1}{2}, & x \geq \frac{1}{2} \\ 0 & x < \frac{1}{2} \end{cases}$$

$$g(x) = \begin{cases} \frac{1}{2} - x, & x \leq \frac{1}{2} \\ 0 & x > \frac{1}{2} \end{cases}$$

then $f \neq 0 \neq g$ and $fg = 0 = gf$.

Group rings

Let R be a ring and G a group (written multiplicatively). The *group ring* $R[G]$ of G over R consists by definition of all formal expressions

$$\sum_{g \in G} a_g \cdot g$$

with $a_g \in R$ for all $g \in G$ and $a_g = 0$ for all except at most finitely many $g \in G$. Two such formal expressions $\sum_{g \in G} a_g \cdot g$ and $\sum_{g \in G} b_g \cdot g$ are considered equal only if $\forall g \in G : a_g = b_g$. Addition is done component wise:

$$\left(\sum_{g \in G} a_g \cdot g \right) + \left(\sum_{g \in G} b_g \cdot g \right) = \sum_{g \in G} (a_g + b_g) \cdot g,$$

and the multiplication one finds by combining the multiplication in R with the one in G:

$$(a_g \cdot g)\cdot(b_h \cdot h) = (a_g b_h)\cdot gh \qquad (a_g, b_h \in R,\ g,h \in G).$$

In the general case (anticipating distributivity) this leads to

$$(\sum_{g\in G} a_g g)\cdot(\sum_{h\in G} b_h h) = \sum_{k\in G}(\sum_{g,h,gh=k} a_g b_h)k.$$

We leave the somewhat elaborate verification that in this way $R[G]$ indeed is a ring as an exercise to the reader.

In case R and G are both commutative, so is $R[G]$. If R contains a unit element 1, then $R[G]$ also contains a unit element, namely $1\cdot e$ where e is the neutral element of G. In what follows we simply denote this unit element of $R[G]$ by 1.

If R is a ring with 1, then G can be considered as a subgroup of $R[G]^\times$ via

$$g = \sum_{h\in G} a_h h, \quad \text{with} \quad a_h = \begin{cases} 0 & \text{if} \quad h \neq g \\ 1 & \text{if} \quad h = g. \end{cases}$$

In case $g \in G$ has order n, with $1 < n < \infty$ then

$$1 + g + g^2 + \ldots + g^{n-1}$$

is a zero divisor of $R[G]$ because

$$(1-g)(1+g+\ldots+g^{n-1}) = 1 - g^n = 0, \quad \text{and } 1-g \neq 0.$$

It is an unsolved problem if $R[G]$ can have zero divisors in the special case that R is a field and G a group with as only element of finite order the neutral element e.

I.3.4 Example. Take $R = \mathbb{Z}$ and $G = (\mathbb{Z}/5\mathbb{Z})^\times = \{\overline{1}, \overline{2}, \overline{3}, \overline{4}\}$. In $R[G]$ we have

$$\begin{gathered}(a\cdot\overline{1} + b\cdot\overline{2} + c\cdot\overline{3} + d\cdot\overline{4})\cdot(k\cdot\overline{1} + \ell\cdot\overline{2} + m\cdot\overline{3} + n\cdot\overline{4}) \\ = \\ (ak + bm + c\ell + dn)\cdot\overline{1} + (a\ell + bk + cn + dm)\cdot\overline{2} \\ + \\ (am + bn + ck + d\ell)\cdot\overline{3} + (an + b\ell + cm + dk)\cdot\overline{4}.\end{gathered}$$

I.4 Exercises

1. Suppose that an element $1'$ in a ring R has the property $1'a = a1' = a$ for all $a \in R$. Prove that $1' = 1$.
2. Let R be a ring. Prove that every $a \in R^\times$ has exactly one inverse.
3. Prove that the set of all even integers $2\mathbb{Z}$ equipped with the familiar addition and multiplication is a commutative ring *without* unit element.
4. Let $\mathrm{M}(2, 2\mathbb{Z})$ be the subset of $\mathrm{M}(2, \mathbb{R})$ (see I.1.4) consisting of all 2×2-matrices having coefficients in $2\mathbb{Z}$. Show that $\mathrm{M}(2, 2\mathbb{Z})$ equipped with the usual matrix addition and matrix multiplication is a *non*-commutative ring *without* unit element.
5. Let $(A, +, 0)$ be an abelian group and define a multiplication on A by $a \cdot b = 0$ for all $a, b \in A$. Prove that A in this way becomes a commutative ring. Does this ring have a unit element?
6. Let R be a non-unitary ring with the property $R^+ \cong \mathbb{Q}/\mathbb{Z}$. Prove that $ab = 0$ for all $a, b \in R$.
7. Let m be an integer that is not a square and put $\alpha := \frac{1+\sqrt{m}}{2} \in \mathbb{C}$.
   (a) For which m is $\mathbb{Z}[\alpha] := \{a + b\alpha : a, b \in \mathbb{Z}\}$ a subring of $\mathbb{C}$?
   (b) Sketch $\mathbb{Z}[\alpha]$ as a subset of the complex plane in case $m = -3$.
8. Let R be a (not necessarily unitary) ring and define an addition and a multiplication on $\mathbb{Z} \times R$ by
$$(n, r) + (m, s) = (n + m, r + s),$$
$$(n, r) \cdot (m, s) = (nm, ns + mr + rs)$$
for $n, m \in \mathbb{Z}, r, s, \in R$ (here
$$ns = s + s + \ldots + s \qquad (n \text{ times})$$
in case $n > 0$, etc.).
   (a) Show that $\mathbb{Z} \times R$ is in this way a ring with 1.
   (b) Prove that every ring can be embedded as a subring in a ring with 1.
9. Let R be a ring with 1 and H an additive subgroup of R. Define $R_0 \subset R$ by $R_0 = \{x \in R : \forall h \in H : xh \in H\}$. Prove that R_0 is a subring of R and $R_0 \neq \{0\}$ in case $R \neq \{0\}$.
10. Suppose R is a ring and $a \in R$. Define $\lambda_a, \rho_a : R \to R$ by $\lambda_a(x) = ax, \rho_a(x) = xa$. Show that λ_a and ρ_a are endomorphisms of the additive group R^+ of R.
11. Let R be a ring. On R we define a new multiplication $*$ by $a * b = ba$, for $a, b \in R$. Show that R with its original addition and this new multiplication is a ring. This ring is called the *opposite* ring of R, notation: R^0.
12. Suppose R is a ring. The *center* of R is
$$\mathcal{Z}(R) = \{a \in R : \forall x \in R : ax = xa\}.$$
Prove that $\mathcal{Z}(R)$ is a subring of R.
13. Suppose that the ring R has the property $x^3 = x$ for all $x \in R$. Show that every $x \in R$ satisfies $x + x + x + x + x + x = 0$.
14. Suppose R is a ring consisting of 10 elements. Prove that R is commutative.
15. (Newton's *Binomium*). Let R be a ring. For $n \in \mathbb{Z}, r \in R$ we define $nr \in R$ as in Exercise 8.

(a) Suppose R is commutative. Prove that

$$(*) \qquad (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^k b^{n-k}$$

for all $a, b \in R$ and $n \in \mathbb{Z}_{>0}$.

(b) Vice versa, prove that if $(*)$ holds for all $a, b \in R$ and $n \in \mathbb{Z}_{>0}$, then the ring R is commutative.

16. Let $\alpha = 1.3247\cdots$ be the real number satisfying $\alpha^3 = \alpha + 1$. Show that $\mathbb{Z}[\alpha]$ defined as $\mathbb{Z}[\alpha] := \{a + b\alpha + c\alpha^2 : a, b, c \in \mathbb{Z}\}$ is a subring of $\mathbb{R}$. Moreover, show that $\alpha, \alpha - 1, \alpha^2 - 1, \alpha^3 - 1 \in \mathbb{Z}[\alpha]^\times$.
17. Let R be a commutative ring with 1 and $n \in \mathbb{Z}_{>0}$. For $A \in \mathrm{M}(n, R)$ one defines $\det(A)$ using the well known formula from linear algebra:

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} \quad \text{with } A = (a_{ij})_{1 \le i,j \le n},$$

where S_n denotes the group of permutations of the set $\{1, \ldots, n\}$ and $\epsilon(\sigma)$ is the sign of $\sigma \in S_n$. Prove $A \in \mathrm{M}(n, R)^\times \iff \det(A) \in R^\times$.

18. Let R be a ring with $1 \neq 0$ and put $T = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(2, R) : c = 0 \right\}$.

(a) Show that T is a subring of $\mathrm{M}(2, R)$ and T is not commutative.

(b) Prove: $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \in T^\times \Leftrightarrow a \in R^\times$ and $d \in R^\times$.

(c) Prove: $T^\times$ is commutative $\iff R^\times = \{1\}$.

(d) In the special case that $R = \mathbb{Z}/2\mathbb{Z}$, show that T is a non-commutative ring with a commutative group of units.

19. Suppose $m \in \mathbb{Z}_{>0}$ satisfies $\sqrt{m} \notin \mathbb{Z}$.

(a) Let $\epsilon = a + b\sqrt{m} \in \mathbb{Z}[\sqrt{m}]^\times$. Prove: $\{\epsilon, \epsilon^{-1}, -\epsilon, -\epsilon^{-1}\} = \{\pm a \pm b\sqrt{m}\}$, and conclude from this: $\epsilon > 1 \Leftrightarrow a > 0 \wedge b > 0$.

(b) Assume that $\mathbb{Z}[\sqrt{m}]^\times \neq \{\pm 1\}$. Show that $\mathbb{Z}[\sqrt{m}]$ contains a smallest unit ϵ_1 with $\epsilon_1 > 1$, and prove that $\mathbb{Z}[\sqrt{m}]^\times = \{\pm \epsilon_1^m : m \in \mathbb{Z}\} \cong (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}$.

(c) Find the smallest unit > 1 in the ring $\mathbb{Z}[\sqrt{5}]$.

20. Let R be a ring and $a \in R$. Define

$$S = \{x \in R : ax = xa\}.$$

(a) Show that S is a subring of R.

(b) Prove that $S^\times = R^\times \cap S$.

21. Let $A \in \mathrm{M}(n, \mathbb{R})$. Prove: A is a left zero divisor $\Leftrightarrow$ A is a right zero divisor $\Leftrightarrow A \neq 0$ and $\det(A) = 0$.
22. Find an example of a commutative ring R with 1 containing an element a with the properties: $a \neq 0$, a is not a unit in R, and a is not a zero divisor of R.
23. Find an example of an infinite commutative ring containing zero divisors.
24. Let K be a field, and define on $R = K \times K$ an addition and a multiplication by

$$(x, y) + (u, v) = (x + u, y + v),$$

$$(x, y) \cdot (u, v) = (xu, xv).$$

(a) Prove that R is a non-commutative ring without a unit element.

(b) Determine the left zero divisors of R and also the right zero divisors of R.

25. Let R be a commutative ring with 1, and R' a subring of R such that $1 \in R'$. For each of the following assertions, provide a proof or a counter example:
    (a) if R is a field, then R' is also a field;
    (b) if R is a domain, then R' is also a domain;
    (c) if R' is a domain, then R is also a domain.
26. Let R_1 and R_2 be rings. Show: $R_1 \times R_2$ is a domain $\Leftrightarrow$ one of the rings R_1, R_2 is a domain and the other is the zero ring $\{0\}$. Do the same exercise with the term 'domain' replaced by 'division ring', and also by 'field '.
27. An *arithmetical function* is a function $f : \mathbb{Z}_{>0} \to \mathbb{C}$. The *sum* $f_1 + f_2$ of two arithmetical functions f_1 and f_2 is defined by
$$(f_1 + f_2)(n) = f_1(n) + f_2(n), \quad \text{for } n \in \mathbb{Z}_{>0}.$$
The *convolution product* $f_1 * f_2$ of two arithmetical functions f_1 and f_2 is defined by
$$(f_1 * f_2)(n) = \sum_{d|n} f_1(d) f_2(\frac{n}{d}) \quad \text{for } n \in \mathbb{Z}_{>0};$$
here the sum is taken over the positive divisors d of n.
    (a) Show that the set R of all arithmetical functions is a *domain* with respect to these two operations.
    (b) Let $f \in R$. Prove: $f \in R^\times \Leftrightarrow f(1) \neq 0$.
28. Let R be a commutative ring, and $S \subset R$ a nonempty subset with the property
$$s, t \in S \implies st \in S.$$
    (a) Show that the relation $\sim$ defined by
$$(a, s) \sim (b, t) \iff \exists u \in S : atu = bsu$$
    is an equivalence relation on $R \times S$.
    (b) Let $S^{-1}R = (R \times S)/\sim$, and let $\frac{a}{s} \in S^{-1}R$ denote the equivalence class containing (a, s). Show that $S^{-1}R$ equipped with the following addition and multiplication is a commutative ring with 1:
$$\frac{a}{s} + \frac{b}{t} = \frac{at + bs}{st}, \qquad \frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st}.$$
    (c) Prove: $S^{-1}R$ is the zero ring $\iff 0 \in S$.
29. Let A be an abelian group. Prove: $\mathrm{End}(A)^\times = \mathrm{Aut}(A)$.
30. Show that $\{f \in C[0,1]$: f is three times continuously differentiable$\}$ is a subring of $C([0,1])$.
31. Let R be a ring with 1 and suppose $a, b \in R$ satisfy $ab = 0$. Show that $(ba)^2 = 0$ and $1 + ba \in R^\times$.
32. (G. Higman, Proc. London Math. Soc. **46** (1940), 231-248).
    (a) Let $a = (13)\cdot\{1-(12)\}, b = 1+(12) \in \mathbb{Z}[S_3]$. Show $ab = 0$, and use Exercise 31 to find a unit of $\mathbb{Z}[S_3]$ not of the form $\pm\sigma$ with $\sigma \in S_3$.
    (b) Suppose G is a group and $g \in G$ has finite order and $< g >$ is not a normal subgroup of G. Prove that $\mathbb{Z}[G]$ contains a unit not of the form $\pm h$ with $h \in G$.
    (c) Let G be a group and suppose $g \in G$ has order 5. Show $1 - g - g^{-1} \in \mathbb{Z}[G]^\times$.
33. A *Boolean ring* (named after the English mathematician George Boole, (1815-1864)) is a ring R such that $x^2 = x$ for all $x \in R$.

(a) Show that $x + x = 0$ for all x in a Boolean ring R.

(b) Prove that every Boolean ring is commutative.

(c) Suppose that the Boolean ring R is a field. Prove that R contains only two elements.

34. Let X be a set and $R = \mathscr{P}(X)$ the set of all subsets of X. For $A, B \in R$ (so $A, B \subset X$) define

$$A + B = (A \cup B) - (A \cap B), \qquad AB = A \cap B.$$

Show that in this way R defines a commutative ring with 1, and R is a field if and only if $\#X = 1$. Moreover, show that R is a *Boolean* ring (Exercise 33).

35. Suppose R is a unitary ring. Let $v \in R$ be a right inverse of $u \in R$: $uv = 1$. Show that the following 3 assertions are equivalent:

(a) u has more than one right inverse;

(b) u is not a unit;

(c) u is a left zero divisor.

36. (Due to Irving Kaplansky, Canadian/American mathematician, 1917–2006.) Let R be a unitary ring and $u \in R$ such that u has more than one right inverse. Prove that u has ∞ many right inverses.
(Hint: if $uv = 1$ and $vu \neq 1$, consider the right inverses $v + (1 - vu)u^n$.)

37. Let R be a finite unitary ring and take $u \in R - \{0\}$. Show that the following assertions are equivalent:

(a) u has a right inverse;

(b) u has a left inverse;

(c) u is not a left zero divisor;

(d) u is not a right zero divisor;

(e) u is a unit.

38. Let R be a unitary ring. Show for $a, b \in R$ that

$$1 - ab \in R^\times \iff 1 - ba \in R^\times \iff \begin{pmatrix} 1 & a \\ b & 1 \end{pmatrix} \in \mathrm{M}(2,R)^\times.$$

II RING HOMOMORPHISMS AND IDEALS

II.1 Ring homomorphisms

In Linear Algebra one especially studies the maps between vector spaces over a given field which preserve the vector space structure; the so-called *linear maps*. In Group Theory one similarly studies *group homomorphisms*: the maps between groups preserving the group structure. We will now do the analogous concept for rings.

II.1.1 Definition. A map $f : R_1 \to R_2$ from a (unitary) ring R_1 to a (unitary) ring R_2 is called a (unitary) *ring homomorphism* if

$$\begin{array}{lrcl} \text{(i)} & f(1) & = & 1, \\ \text{(ii)} & f(a+b) & = & f(a)+f(b), \\ \text{(iii)} & f(ab) & = & f(a)\cdot f(b) \end{array}$$

for all $a, b \in R_1$. (The '(unitary)' in the definition of ring homomorphism corresponds to the first condition (i).)

A bijective ring homomorphism is called a *ring isomorphism*, the inverse of such a bijection is ring homomorphism as well. Two rings R_1 and R_2 are called *isomorphic* if a ring isomorphism $R_1 \to R_2$ exists; notation: $R_1 \cong R_2$. An isomorphism from a ring R to itself is called a (*ring*)-*automorphism* of R.

A unitary ring homomorphism from a field to a field is called a *field homomorphism*, and similarly we have a *field isomorphism* and a *field automorphism*.

II.1.2 Examples. 1. If R' is a subring of a ring R then the inclusion map $R' \to R$ is an injective ring homomorphism.

2. Let $n \in \mathbb{Z}_{>0}$. The canonical map

$$f : \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}, \qquad f(a) = \overline{a},$$

is a ring homomorphism, because $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{a} \cdot \overline{b} = \overline{ab}$.

3. For every $s \in R^\times$ the map (conjugation by s):

$$\gamma_s : R \longrightarrow R, \qquad r \mapsto srs^{-1}$$

is a (bijective) ring homomorphism. If R is commutative, then evidently $\gamma_s = id_R$ for all $s \in R^\times$. In case $R = M(n, \mathbb{R})$ a change of basis in $\mathbb{R}^n$ induces a conjugation γ_s on $M(n, \mathbb{R})$.

4. If R_1, R_2 are rings then the projection $f : R_1 \times R_2 \to R_1$, $f((a,b)) = a$ is a ring homomorphism.

II.1.3 Definition. If $f : R_1 \to R_2$ is a ring homomorphism, then the *image* of f is defined as

$$f(R_1) := \{y \in R_2 : \exists x \in R_1 \text{ with } y = f(x)\}.$$

The *kernel* of f is (as in the case of additive groups) defined as:

$$\mathrm{Ker}(f) := \{x \in R_1 : \ f(x) = 0\}.$$

Ring homomorphisms have properties that in many ways are analogous to those of group homomorphisms. For example: is $f : R_1 \to R_2$ a ring homomorphism then the image $f(R_1)$ of f is a *subring* of R_2. The easy proof of this is left as an exercise.

Since a ring homomorphism is in particular a homomorphism of additive groups, one finds:

$$\mathrm{Ker}(f) = \{0\} \quad \Longleftrightarrow \quad f \quad \text{is injective.}$$

The kernel of a (unitary) ring homomorphism f is a possibly non-unitary subring, namely $f(1)$ is not necessarily equal to 0.

II.2 Ideals

In Group Theory it turned out that not *all* subgroups occur as kernels of group homomorphisms: only the *normal subgroups* do. Similarly we will now see that not all non-unitary subrings occur as kernel of some ring homomorphism: only so-called *ideals* do, which we will now define.

II.2.1 Definition. Let R be a ring. An *ideal* of R is a subset $I \subset R$ satisfying:

(I1) I is a subgroup of the additive group of R, in other words:

- **(H0)** $0 \in I$;
- **(H1)** $a - b \in I$ for all $a, b \in I$;

(I2) for all $r \in R$ and $a \in I$ we have $ra \in I$ and $ar \in I$.

II.2.2 Remark. Instead of 'ideal' one also says 'two sided ideal'. Replacing (I2) by the weaker assumption

(I2') $\forall r \in R : \forall a \in I : ra \in I$

the definition of a *left ideal* of R is obtained. Taking ar instead of ra one has the definition of a *right ideal*.

An example of a left ideal which is not a right ideal - hence not an ideal - is given in Exercise II.5.23. We will mostly be interested in commutative rings, and here the three notions evidently coincide.

II.2.3 Example. Trivial examples of ideals are $\{0\}$ and R itself. For every $n \in \mathbb{Z}$ the subset

$$n\mathbb{Z} := \{nk \in \mathbb{Z} : \ k \in \mathbb{Z}\} \qquad (\subseteq \mathbb{Z})$$

of $\mathbb{Z}$ is an ideal of $\mathbb{Z}$ (verify!). ∎

II.2.4 Remark. An ideal is a subring, in general non-unitary, however the converse (is every subring an ideal?) is far from true in general: $\mathbb{Z}$ is a subring of $\mathbb{Q}$ but not an ideal of $\mathbb{Q}$ since, e.g.,

$$r = \frac{1}{2} \in \mathbb{Q}, \quad a = 1 \in \mathbb{Z}, \quad \text{but} \quad ra = \frac{1}{2} \notin \mathbb{Z}$$

so (I2) does not hold. In general: is R a ring with 1 and I an ideal of R with $1 \in I$, then $I = R$ (namely, apply (I2) to $a = 1$); see II.4.4 for a generalisation of this.

II.2.5 Theorem. *If $f : R_1 \to R_2$ is a ring homomorphism,* $\mathrm{Ker}(f)$ *is an ideal of* R_1.

Proof. We verify (I1) and (I2) for $I = \mathrm{Ker}(f)$.
(I1) This follows from the fact that f is a group homomorphism $R_1^+ \longrightarrow R_2^+$.
(I2) For $r \in R_1$, $a \in \mathrm{Ker}(f)$ one has $f(a) = 0$, hence

$$f(ra) = f(r)f(a) = f(r)\cdot 0 = 0, \qquad f(ar) = f(a)f(r) = 0 \cdot f(r) = 0,$$

which shows $ra, ar \in \mathrm{Ker}(f)$, as desired. This shows II.2.5. ∎

Later (see II.3.5) we will see that the converse of II.2.5 holds as well: every ideal $I \subset R$ is the kernel of a suitably chosen ring homomorphism.

II.2.6 Example. For $n > 1$ the kernel of the canonical ring homomorphism

$$f : \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}, \qquad a \mapsto \bar{a}$$

equals the ideal $n\mathbb{Z}$ discussed in Example II.2.3. ∎

II.2.7 Example. A straightforward verification shows that

$$f : \mathbb{Z}[i] \longrightarrow \mathbb{F}_2 (= \mathbb{Z}/2\mathbb{Z}), \qquad a + bi \mapsto \bar{a} + \bar{b} \qquad (a, b \in \mathbb{Z})$$

is a (surjective, unitary) ring homomorphism. We claim:

$$\mathrm{Ker}(f) = \{2r + (1+i)s \in \mathbb{Z}[i] : r, s \in \mathbb{Z}[i]\} = \{(1+i)t \in \mathbb{Z}[i] : t \in \mathbb{Z}[i]\}.$$

(If everywhere in this example $i = \sqrt{-1}$ is replaced by $\sqrt{-5}$, then the first '=' still holds, but the second one does not: see Exercise II.5.17.) We start by proving the first equality:
'$\supset$' since f is a ring homomorphism, all $r, s \in \mathbb{Z}[i]$ satisfy

$$f(2r + (1+i)s) = f(2)f(r) + f(1+i)f(s) = 0 \cdot f(r) + 0 \cdot f(s) = 0.$$

'$\subset$' If $a + bi \in \mathrm{Ker}(f)$ for some $a, b \in \mathbb{Z}$, then $a + b \equiv 0 \bmod 2$ hence $a = b + 2k$ for some $k \in \mathbb{Z}$. Then $a + bi = b + 2k + bi = 2k + (1+i)b$ with $k, b \in \mathbb{Z} \subset \mathbb{Z}[i]$.

For the second '=' sign one observes:

$$2r + (1+i)s = (1+i)(1-i)r + (1+i)s = (1+i)\cdot((1-i)r + s) = (1+i)t,$$

for $t = (1-i)r + s$, showing '$\subset$'. Moreover '$\supset$' is evident since one can take $r = 0$, $s = t$.

By Theorem II.2.5 both sets are ideals. Check that this can also be seen directly from the definition of an ideal. ∎

As we saw in the example above, the subsets

$$2\mathbb{Z}[i] + (1+i)\mathbb{Z}[i], \qquad (1+i)\mathbb{Z}[i]$$

are ideals in $\mathbb{Z}[i]$, in fact the same ideals. More generally:

II.2.8 Definition. Let R be a commutative, unitary ring and take $a_1, a_2 \ldots, a_n \in R$. The ideal *generated* by $a_1, a_2, \ldots, a_n$ is defined as:

$$Ra_1 + Ra_2 + \ldots + Ra_n = \{r_1a_1 + r_2a_2 + \ldots + r_na_n : r_1, r_2, \ldots, r_n \in R\}.$$

If it is clear from the context which ring R is considered, then one writes

$$(a_1, a_2, \ldots, a_n) := Ra_1 + Ra_2 + \ldots + Ra_n.$$

An ideal $I \subset R$ is called a *principal ideal* if $I = (a) = Ra$ for some $a \in R$.

Verify using Definition II.2.1 that indeed this defines an ideal (in case R is not commutative, one in general only obtains a left ideal in this way).

II.2.9 Examples. The ideal $(2, 1+i) \subset \mathbb{Z}[i]$ is principal, namely we showed earlier that $(2, 1+i) = (1+i)$.

Also the ideal $I = (4, 6) \subset \mathbb{Z}$ turns out to be a principal ideal: $2 = (-1)4 + 6 \in I$ hence $2\mathbb{Z} \subset I$ (use (I2)) and on the other hand $4, 6 \in 2\mathbb{Z}$ which implies $I = 4\mathbb{Z}+6\mathbb{Z} \subset 2\mathbb{Z}$. This shows $(4, 6) = (2)$.

Note that in the above definition, since R is unitary, $a_1, a_2, \ldots, a_t$ are contained in the ideal $Ra_1 + Ra_2 + \ldots + Ra_t$. Any ideal I containing all a_i also contains by (I2) all elements in $Ra_1, Ra_2, \ldots, Ra_t$ and therefore using (I1) all elements from $Ra_1+Ra_2+\ldots+Ra_t$. Hence $Ra_1+\ldots+Ra_t$ is the *smallest* ideal containing $a_1, a_2, \ldots, a_t$.

In Section II.4 we will study generators of ideals in more detail.

Given any ring R one can define (see Chapter III.1 below) the ring of polynomials in the variable X with coefficients in R, denoted $R[X]$. This is analogous to probably familiar polynomial rings such as $\mathbb{Z}[X]$ and $\mathbb{R}[X]$. For this reason the next result is given in a rather general setting.

II.2.10 Theorem. *Let R be a commutative ring with 1 and $\alpha \in R$.*

The map 'evaluation in α' defined as

$$\mathrm{ev}_\alpha : R[X] \longrightarrow R, \qquad \mathrm{ev}_\alpha\left(\sum_{i=0}^{n} a_i X^i\right) = \sum_{i=0}^{n} a_i \cdot \alpha^i$$

is a (surjective, unitary) ring homomorphism (note that $\mathrm{ev}_\alpha(f) = f(\alpha)$).

Moreover

$$\mathrm{Ker}(\mathrm{ev}_\alpha) = (X - \alpha) = \{(X - \alpha)g : \ g \in R[X]\}.$$

Proof. The fact that ev_α defines a ring homomorphism we will show in much more generality in Theorem III.2.1). It is surjective since any $r \in R$ satisfies $r = \mathrm{ev}_\alpha(r)$. We now show the second assertion.

'$\supset$': Note that $\mathrm{ev}_\alpha(X - \alpha) = \alpha - \alpha = 0$, hence $X - \alpha \in \mathrm{Ker}(\mathrm{ev}_\alpha)$, and because $\mathrm{Ker}(\Phi_\alpha)$ is an ideal then also $R[X](X - \alpha) \subset \mathrm{Ker}(\mathrm{ev}_\alpha)$.

'$\subset$': Suppose that $\sum\limits_{i=0}^{n} a_i X^i \in \mathrm{Ker}(\mathrm{ev}_\alpha)$, then $\sum\limits_{i=0}^{n} a_i \alpha^i = 0$, hence

$$\begin{array}{rl} \sum_{i=0}^{n} a_i X^i & = \sum_{i=0}^{n} a_i X^i - \sum_{i=0}^{n} a_i \alpha^i \\ = & \sum_{i=0}^{n} a_i (X^i - \alpha^i) \\ = & \sum_{i=0}^{n} a_i (X^{i-1} + \alpha X^{i-2} + \ldots + \alpha^{i-3} X^2 + \alpha^{i-2} X + \alpha^{i-1})(X - \alpha) \\ \in & R[X](X - \alpha). \end{array}$$

This proves Theorem II.2.10. ∎

II.2.11 Examples. A simple special case is the ring homomorphism

$$\mathrm{ev}_0 : \mathbb{R}[X] \longrightarrow \mathbb{R}, \qquad f \mapsto f(0).$$

Writing $f = \sum a_i X^i$ we find $f(0) = a_0$ hence

$$\mathrm{Ker}(\mathrm{ev}_0) = \{f \in \mathbb{R}[X] : \ f = \sum a_i X^i \text{ and } a_0 = 0\}.$$

Since $a_0 = 0$ precisely when $f = Xg$ with $g = \sum_{i=1}^{n} a_i X^{i-1} \in \mathbb{R}[X]$, one concludes that indeed $\mathrm{Ker}(\mathrm{ev}_0) = X\mathbb{R}[X]$.

For the next example, note that in $\mathbb{R}[X,Y]$ every polynomial can be written as

$$\sum_{i,j} a_{ij}X^iY^j = \sum_{j=0}^{m}\left(\sum_{i=0}^{n} a_{ij}X^i\right)Y^j = \sum_{j=0}^{m} f_j(X)Y^j$$

where $f_j(X) = \sum_{i=0}^{n} a_{ij}X^i$. Hence a polynomial in two variables X, Y can be regarded as a polynomial in one variable Y with coefficients from the ring $\mathbb{R}[X]$:

$$\mathbb{R}[X,Y] = (\mathbb{R}[X])[Y].$$

Any $f \in \mathbb{R}[X]$ therefore defines a ring homomorphism

$$\mathrm{ev}_f : \mathbb{R}[X,Y] = (\mathbb{R}[X])[Y] \longrightarrow \mathbb{R}[X], \qquad F(X,Y) \mapsto F(X, f(X)).$$

The kernel of this ring homomorphism is by Theorem II.2.10 the ideal

$$\mathrm{Ker}(\mathrm{ev}_f) = \{(Y - f(X))G(X,Y) : \ G(X,Y) \in \mathbb{R}[X,Y]\}.$$

A special case is obtained when $f = 0$. Verify (without using the theorem) that in this case indeed $(Y) \subset \mathbb{R}[X,Y]$ is the kernel of ev_0.

II.3 The factor ring R/I.

Let R be a ring and $I \subset R$ an ideal. Then I is a *normal subgroup* of the additive group of R (by (I1) and the fact that R^+ is abelian). The set of residue classes

$$R/I := \{\overline{a} := a + I \subset R : \ a \in R\}$$

of I in R is therefore an additive *group*, with addition $\overline{a} + \overline{b} = \overline{a+b}$. Two elements $\bar{a}, \bar{b} \in R/I$, so two subsets of R as above are equal precisely when $a - b \in I$:

$$\bar{a} = \bar{b} \quad \Longleftrightarrow \quad a + I = b + I \quad \Longleftrightarrow \quad a - b \in I,$$

namely $\Rightarrow$: $a+I = b+I$ and $0 \in I$ implies $a+0 = b+i$ for some $i \in I$, hence $a-b = i \in I$; $\Leftarrow$: If $a = b + i$ with $i \in I$ then it follows, since I is a subgroup of R^+, that $i + I = I$ and therefore $a + I = b + i + I = b + I$.

We now define a *multiplication* on R/I by

$$(a + I)\cdot(b + I) := ab + I, \qquad \text{i.e.,} \quad \overline{a}\cdot\overline{b} = \overline{ab}.$$

To verify that this is well one needs to show: if

$$\overline{a} = \overline{a}_1, \ \overline{b} = \overline{b}_1 \qquad \text{then} \quad \overline{ab} = \overline{a_1 b_1}.$$

The condition $\overline{a} = \overline{a}_1$ implies $a_1 = a + i$ with $i \in I$, and analogously $b_1 = b + j$ with $j \in I$. Hence:

$$a_1 b_1 = (a + i)(b + j) = ab + ib + aj + ij = ab + k \quad \text{where } k \in I,$$

because I is an ideal which implies ai, bj, $ij \in I$ by (I2), and then by (I1) the sum k is in I as well. Now $a_1 b_1 = ab + k$, $k \in I$, is equivalent to $\overline{a_1 b_1} = \overline{ab}$. As a result, the multiplication on R/I is well defined.

II.3.1 Remark. The rule $\overline{a}\,\overline{b} = \overline{ab}$ is *not* well defined in the group $\mathbb{Q}/\mathbb{Z}$ (try for example $a = \frac{1}{2}$, $b = 3$, $a_1 = \frac{1}{2}$, $b_1 = 2$). The subgroup $\mathbb{Z}$ of $\mathbb{Q}$ is therefore not an ideal in $\mathbb{Q}$, as we already noticed in II.2.4.

It turns out that $(R/I, +, \cdot, \overline{0}, \overline{1})$ is a *ring*, called the factor ring (or quotient ring) of R modulo I.

By way of example we verify one of the distributive laws (R3):

$$\begin{aligned}
\overline{a}(\overline{b}+\overline{c}) &= \overline{a}\,\overline{(b+c)} && \text{(by definition of } +) \\
&= \overline{a(b+c)} && \text{(by definition of } \cdot) \\
&= \overline{ab+ac} && \text{(since (R3) holds in } R) \\
&= \overline{ab}+\overline{ac} && \text{(by definition of } +) \\
&= \overline{a}\cdot\overline{b}+\overline{a}\cdot\overline{c} && \text{(by definition of } \cdot).
\end{aligned}$$

In a similar way one verifies the remaining ring-axioms.

If R is commutative, so is R/I. If R contains a 1 then $\overline{1}$ is a unit element of R/I.

II.3.2 Example. The rings $\mathbb{Z}/n\mathbb{Z}$ are special cases of this construction. Taking for example $n = 6$ we see that R/I may have zero divisors even if R does not. ▬■

II.3.3 Definition. Given a ring R and an ideal $I \subset R$, the map

$$\phi : R \longrightarrow R/I, \quad \phi(a) := \overline{a} = a + I,$$

is called the *natural* or *canonical* map.

II.3.4 Theorem. *For any ring R and ideal $I \subset R$ the natural map $\phi : R \to R/I$ is a surjective ring homomorphism with kernel* $\mathrm{Ker}(\phi) = I$.

Proof. Surjectivity of ϕ is clear. From

$$\phi(a+b) = \overline{a+b} = \overline{a}+\overline{b} = \phi(a)+\phi(b),$$

$$\phi(ab) = \overline{ab} = \overline{a}\cdot\overline{b} = \phi(a)\cdot\phi(b)$$

it follows that ϕ is a ring homomorphism. Finally

$$\phi(a) = \overline{0} \iff \overline{a} = \overline{0} \iff a \in I$$

hence $I = \mathrm{Ker}(\phi)$. This proves Theorem II.3.4. ■

II.3.5 Corollary. *Let R be a ring and $I \subset R$. Then:*

I is an ideal in R $\Leftrightarrow$ a ring homomorphism $f : R \to R'$ exists with $\mathrm{Ker}(f) = I$.

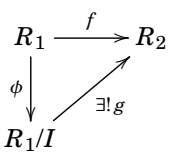
Proof. $\Leftarrow$: This is II.2.5. $\Rightarrow$: take $R' = R/I, f = \phi$ as in II.3.4. This shows II.3.5. ■

The preceding results are analogs of results in Group Theory. We will now formulate the results corresponding to the homomorphism- and isomorphism theorems. Because of the far reaching analogy the proofs will be quite brief.

II.3.6 Theorem. (The homomorphism theorem for rings). *Let $f : R_1 \to R_2$ be a ring homomorphism and $I \subset R_1$ an ideal with $I \subset \mathrm{Ker}(f)$. Take $\phi : R_1 \to R_1/I$ the canonical ring homomorphism.*

Then a unique ring homomorphism $g : R_1/I \to R_2$ exists with $f = g \circ \phi$. Moreover,

$$\mathrm{Ker}(g) = \phi(\mathrm{Ker}(f)),$$



Proof. If g exists, then $f(a) = g(\phi(a)) = g(a+I)$ for all $a \in R_1$. We would therefore like to define: $g(a+I) := f(a)$. The (possible) problem with this is that possibly $a+I = b+I$ whereas $f(a) \neq f(b)$. In this case the proposed assignment does not define the image of a residue class, since the choice of a representing element affects the outcome.

However, the condition $I \subset \mathrm{Ker}(f)$ implies that the problem described here does not occur:

$$a+I = b+I \quad \Rightarrow \quad a-b \in I \subset \mathrm{Ker}(f) \quad \Rightarrow \quad f(a-b) = 0 \quad \Rightarrow f(a) = f(b).$$

We can therefore assign a uniquely determined element $f(a)$ in R_2 to every residue class $a+I$ in R_1/I. This defines

$$g : R_1/I \longrightarrow R_2, \qquad a+I \mapsto f(a).$$

It is not hard to verify that g is a ring homomorphism with kernel $\phi(\mathrm{Ker}(f))$. ∎

II.3.7 Theorem. (The first isomorphism theorem for rings). *Let $f : R_1 \to R_2$ be a ring homomorphism.*

Then there exists an isomorphism of rings:

$$R_1/\mathrm{Ker}(f) \xrightarrow{\cong} f(R_1), \qquad \bar{a} = a + \mathrm{Ker}(f) \mapsto f(a) \quad (a \in R_1).$$

In particular, in case f is surjective one finds $R_1/\mathrm{Ker}(f) \cong R_2$.

Proof. Applying the previous result with $I = \mathrm{Ker}(f)$ yields a ring homomorphism $g : R_1/\mathrm{Ker}(f) \to R_2$ with $\mathrm{Ker}(g) = \phi(\mathrm{Ker}(f)) = \bar{0}$, hence g is injective. As a consequence, $g : R_1/\mathrm{Ker}(f) \to f(R_1) \subset R_2$ is a bijective ring homomorphism, so an isomorphism of rings. This proves II.3.7. ∎

II.3.8 Example. Combining Example II.2.7 and Theorem II.2.10 and the second example from Examples II.2.11 with the first isomorphism theorem shows:

$$\mathbb{Z}[i]/(1+i) \cong \mathbb{F}_2, \qquad R[X]/(X-\alpha) \cong R, \qquad \mathbb{R}[X,Y]/(Y-f(X)) \cong \mathbb{R}[X].$$

▬■

II.3.9 Example. Let $N \in \mathbb{Z}$ and $(\mathbb{Z}/N\mathbb{Z})[X]$ the ring of polynomials with coefficients in $\mathbb{Z}/N\mathbb{Z}$. We will use the first isomorphism theorem to prove:

$$\mathbb{Z}[X]/N\mathbb{Z}[X] \cong (\mathbb{Z}/N\mathbb{Z})[X].$$

Define the map

$$\psi : \mathbb{Z}[X] \longrightarrow (\mathbb{Z}/N\mathbb{Z})[X], \qquad \sum_{i=0}^{n} a_i X^i \mapsto \sum_{i=0}^{n} \overline{a_i} X^i,$$

with $\overline{a_i} = a_i \bmod N \in \mathbb{Z}/N\mathbb{Z}$. Verify for yourself that ψ is a surjective ring homomorphism. If $\sum_{i=0}^{n} \overline{a_i} X^i = 0$ then $\overline{a_i} = \overline{0} \quad \forall i$ (since the definition of polynomials includes the property $\sum c_i X^i = \sum d_i X^i$ precisely when $a_i = b_i$ for all i). Now $\overline{a_i} = \overline{0}$ means $a_i = N b_i$ for some $b_i \in \mathbb{Z}$. Taking out the factor N one concludes

$$\psi(\sum_{i=0}^{n} a_i X^i) = 0 \Longleftrightarrow \sum_{i=0}^{n} a_i X^i = N(\sum_{i=0}^{n} b_i X^i),$$

hence

$$\mathrm{Ker}(\psi) = N\mathbb{Z}[X] \subset \mathbb{Z}[X].$$

The first isomorphism theorem therefore yields the desired isomorphism. ▬■

For the ring theoretic equivalent of the second isomorphism theorem we refer to Exercise II.5.28. The third isomorphism theorem corresponds to the next result.

II.3.10 Theorem. *Let R be a ring and I an ideal of R and $\phi : R \to R/I$ the natural map. There exists a bijection between the ideals J' of R/I and the ideals J of R containing I. This bijection assigns to the ideal J' of R/I the ideal J of R given by*

$$J := \{x \in R : \ \phi(x) \in J'\} \qquad (\textit{note that} \quad \phi(J) = J').$$

Moreover, for any ideal J of R containing I we have

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

Proof. This is completely analogous to the proof of the corresponding theorem in Group Theory, see Exercise 25 on page 33. ∎

II.3.11 Example. Let $(a, b) \in \mathbb{R}$ and put

$$I := (Y - b) = (Y - b)\mathbb{R}[X, Y]$$

and

$$J := (X - a, Y - b) = (X - a)\mathbb{R}[X, Y] + (Y - b)\mathbb{R}[X, Y].$$

Then I and J are ideals in $\mathbb{R}[X, Y]$ with $I \subset J$. We will show that

$$\mathbb{R}[X, Y]/J \cong \mathbb{R}.$$

The ring $\mathbb{R}[X, Y]/I$ has a simple description, see II.3.8:

$$\Phi_b : \mathbb{R}[X, Y]/I \xrightarrow{\cong} \mathbb{R}[X], \qquad F + I \mapsto F(X, b).$$

Using the ring isomorphism Φ_b it follows that

$$(\mathbb{R}[X, Y]/I) \Big/ \phi(J) \qquad \cong \mathbb{R}[X]/\Phi_b(\phi(J)), \qquad \text{with} \quad \phi : \mathbb{R}[X, Y] \to \mathbb{R}[X, Y]/I$$

the natural map.

In order to apply Theorem II.3.10 we determine the ideal $\Phi_b(\phi(J))$ of $\mathbb{R}[X]$. Define $\Psi_b := \Phi_b \circ \phi : \mathbb{R}[X, Y] \to \mathbb{R}[X]$. Since $\Phi_b(\phi(F)) = \Phi_b(F + I) = F(X, b)$, we find:

$$\Phi_b \circ \phi = \Psi_b : \mathbb{R}[X, Y] \longrightarrow \mathbb{R}[X], \qquad F(X, Y) \mapsto F(X, b).$$

The ideal $\Psi_b(J) = \Phi_b(\phi(J))$ of $\mathbb{R}[X]$ therefore equals

$$\begin{aligned} \Phi_b(\phi(J)) = \Psi_b(J) &= \Psi_b(Y - b)\Psi_b(\mathbb{R}[X, Y]) + \Psi_b(X - a)\Psi_b(\mathbb{R}[X, Y]) \\ &= 0 + (X - a)\mathbb{R}[X] \\ &= (X - a)\mathbb{R}[X]. \end{aligned}$$

We conclude

$$\mathbb{R}[X, Y]/J \cong \mathbb{R}[X]/(X - a) \cong \mathbb{R}, \qquad F + J \mapsto F(X, b) + (X - a) \mapsto F(a, b),$$

again using Theorem II.2.10.

The canonical map $\psi : \mathbb{R}[X, Y] \to \mathbb{R}[X, Y]/J$ being a surjective homomorphism with $\mathrm{Ker}(\psi) = J$ therefore implies that

$$\psi : \mathbb{R}[X, Y] \longrightarrow \mathbb{R}, \qquad F \mapsto F(a, b)$$

is a (surjective) ring homomorphism with kernel J, showing $\mathbb{R}[X.Y]/J \cong \mathbb{R}$. In the special case $a = b = 0$ you may try to find a more direct proof of this. ∎

II.4 Calculating with ideals

In II.2 we defined

$$(a_1,\ldots,a_n) = \{r_1a_1 + \ldots + r_na_n : \ r_i \in R\},$$

the ideal generated by $a_1,\ldots,a_n$ in a commutative ring R with 1. Moreover (Definition II.2.8) an ideal in such a ring R is called principal if it can be generated by a single element.

II.4.1 Definition. We say that a ring R is a *principal ideal ring* if every ideal in R is a principal ideal.

II.4.2 Theorem. *The ring $\mathbb{Z}$ is a principal ideal ring.*

Proof. Let I be an ideal in $\mathbb{Z}$. If $I = \{0\}$ then certainly I is principal. If $I \neq \{0\}$ then an $n \in I$ with $n \neq 0$ exists. Put

$$N := \min\{n \in I : \ n > 0\}.$$

We claim that $I = (N) = N\mathbb{Z}$, a principal ideal.

Since $N \in I$ it follows by (I2) that $N\mathbb{Z} \subset I$. Vice versa let $n \in I$. Division with remainder in $\mathbb{Z}$ yields $q \in \mathbb{Z}$ and $r \in \mathbb{Z}_{\geq 0}$ satisfying

$$n = qN + r \qquad 0 \leq r < N.$$

Since I is an additive group and $n, qN \in I$, it follows that $r = n - qN \in I$. Given that N is the smallest positive element in I and $0 \leq r < N$, we conclude $r = 0$. As a result,

$$n = qN \in N\mathbb{Z} = (N),$$

so $I \subset N\mathbb{Z}$ and therefore $I = N\mathbb{Z}$ is a principal ideal, proving Theorem II.4.2. ∎

II.4.3 Example. We show that the ideal

$$(X,Y) = X\mathbb{R}[X,Y] + Y\mathbb{R}[X,Y] \subset \mathbb{R}[X,Y]$$

is *not* a principal ideal.

Suppose that (X,Y) were a principal ideal. Then $f \in \mathbb{R}[X,Y]$ exists,

$$f = a_{00} + a_{10}X + a_{01}Y + \ldots, \qquad \text{with} \quad f\mathbb{R}[X,Y] = (X,Y).$$

From $f\mathbb{R}[X,Y] \subset X\mathbb{R}[X,Y] + Y\mathbb{R}[X,Y]$ one knows $f = f \cdot 1 = Xh + Yk$ for certain $h, k \in \mathbb{R}[X,Y]$. This implies $a_{00} = 0$. Note

$$(f) \supset (X,Y) \Longrightarrow \begin{cases} X &= fh \\ Y &= fk \end{cases}$$

with $h, k \in \mathbb{R}[X,Y]$. From $X = fh$ one concludes $a_{10} \neq 0$ and $h(0,0) = a_{10}^{-1} \neq 0$. Similarly $Y = fk$ implies $a_{01} \neq 0$ (here one uses $a_{00} = 0$). This yields a contradiction, since:

$$fh = (a_{10}X + a_{01}Y + \ldots)(a_{10}^{-1} + \ldots) = X + a_{01}a_{10}^{-1}Y + \ldots \neq X,$$

using $a_{01}a_{10}^{-1} \neq 0$. Conclusion: (X,Y) is not principal. ∎

We will see later that $\mathbb{R}[X]$ and $\mathbb{Z}[i]$ are principal ideal rings (the proof is analogous to that of Theorem II.4.2), and that $\mathbb{Z}[X]$ and $\mathbb{Z}[\sqrt{-5}]$ are not principal ideal rings.

We now show (Corollary II.4.5) that every division ring is a principal ideal ring.

II.4.4 Theorem. *If R is a unitary ring and I an ideal of R with $I \cap R^\times \neq \varnothing$, then $I = R$.*

Proof. Let $a \in I \cap R^\times$. From $a \in R^\times$ it follows that $\exists b \in R : ab = 1$. Then (I2) (with $r = b$) implies $1 \in I$. Again using (I2) (with $a = 1$) shows that every $r \in R$ is in I, hence $R = I$. This proves Theorem II.4.4. ∎

II.4.5 Corollary. *The only ideals in a division ring R are $\{0\}$ and $R = R \cdot 1$. In particular division rings are principal ideal rings.*

Proof. Let $I \subset R$ be an ideal. If I contains an element $a \neq 0$ then $a \in R^\times$ so $I = R$ by II.4.4. If I does not contain an element $\neq 0$ then $I = \{0\}$. This shows Corollary II.4.5. ∎

II.4.6 Corollary. *Every unitary ring homomorphism $f : D \to R$ from a division ring D to a ring $R \neq \{0\}$ is injective. In particular every field homomorphism is injective.*

Proof. The kernel $\mathrm{Ker}(f)$ of f is an ideal of D, so $\mathrm{Ker}(f) = \{0\}$ or D (by II.4.5). However $f(1) = 1 \neq 0$ since $R \neq \{0\}$, hence $1 \notin \mathrm{Ker}(f)$ which implies $\mathrm{Ker}(f) \neq D$. Therefore $\mathrm{Ker}(f) = \{0\}$ and as a consequence f is injective. The second assertion follows immediately from the first. ∎

II.4.7 Definition. For a ring R and ideals I, J of R one defines the *sum* of I and J by

$$I + J = \{x + y : x \in I, y \in J\}.$$

Using Definition II.2.1 it is immediate that $I + J$ is an ideal of R. Moreover $I + J$ contains the two ideals I and J, and every ideal containing both I and J also contains $I + J$. Hence $I + J$ is the *smallest* ideal containing I and J. This naturally raises the question to describe the *largest* ideal of R contained in the ideals I and J. It turns out that this is the *intersection* $I \cap J$: verify for yourself that indeed this is an ideal of R and that any ideal of R contained in both I and J, is contained in $I \cap J$.

II.4.8 Definition. Ideals I and J in a unitary ring R are called *coprime* or *relative prime* if

$$I + J = R.$$

In Example II.4.10 this terminology will be explained using the special case $R = \mathbb{Z}$. In general, given that R is unitary

$$\begin{aligned} I + J = R &\iff 1 \in I + J && \text{(by II.4.4)} \\ &\iff \exists x \in I,\, y \in J : x + y = 1. \end{aligned}$$

II.4.9 Definition. The *product* of ideals I and J in a ring R is defined by

$$I \cdot J = \left\{ \sum_{i=1}^{n} x_i y_i : \ n \in \mathbb{Z}_{\geq 0},\, x_i \in I,\, y_i \in J \right\}.$$

Again it is not hard to verify that the product of ideals I, J in a ring R is also an ideal of R; note that Exercise 31 on page 33 shows that $\{xy : x \in I, y \in J\}$ is not necessarily an ideal of R. Since $x_i y_i \in I$ for all $x_i \in I$ and $y_i \in J$ (by (I2)), every element $\sum_{i=1}^{n} x_i \cdot y_i$ of $I \cdot J$ is also in I. Similarly $I \cdot J \subset J$ which shows that $I \cdot J \subset I \cap J$.

$$I \cdot J \subset \begin{matrix} \subset I \\ I \cap J \\ \subset J \end{matrix} \qquad \subset I + J \subset R.$$

Sums, intersections, and products can be defined in general for more than two (but in case of products, finitely many) ideals. The result is again an ideal.

II.4.10 Example. We now consider the concepts introduced above in the special case $R = \mathbb{Z}$. Every ideal of $\mathbb{Z}$ is a principal ideal $\mathbb{Z}a$ (see II.4.2).

The *sum* of two ideals (not both $\{0\}$) corresponds in this case to the greatest common divisor of the generators:

$$\mathbb{Z}a + \mathbb{Z}b = \mathbb{Z}d \quad \text{with} \quad d = \gcd(a,b).$$

Proof: By Theorem II.4.2 one has $\mathbb{Z}a + \mathbb{Z}b = \mathbb{Z}N$, i.e., $(a, b) = (N)$ for some $N \in \mathbb{Z}_{>0}$. Hence we wish to show that $d = N$. Since a and b are divisible by d and $k, l \in \mathbb{Z}$ exist with $ak + bl = N$, one concludes $d|N$. On the other hand $(a, b) = (N)$ implies that $a = a \cdot 1 + b \cdot 0 = r_1 N$ and similarly $b = r_2 N$ for certain $r_1, r_2 \in \mathbb{Z}$. Therefore $N|a$ and $N|b$ which shows that N is a common divisor of a and b. Now all common divisors divide the *largest* common divisor, so $N|d$. From $d, N \in \mathbb{Z}_{>0}$, $d|N$, $N|d$ one concludes $d = N$ which finishes the proof.

In particular the argument above shows

$$\gcd(a,b) = d \implies ka + lb = d \quad \text{for certain } k, l \in \mathbb{Z},$$

a statement called 'Bézout's identity' (after the French mathematician Étienne Bézout, 1730–1783) although it was in fact proven much earlier by the French mathematician Claude Gaspard Bachet de Méziriac (1581–1638).

As a consequence, the ideals $\mathbb{Z}a$ and $\mathbb{Z}b$ are relative prime if and only if a, b are coprime, i.e., $\gcd(a,b) = 1$. This explains the terminology introduced in Definition II.4.8.

The *intersection* of two ideals $\mathbb{Z}a, \mathbb{Z}b$ corresponds to taking the *least common multiple* of the generators:

$$\mathbb{Z}a \cap \mathbb{Z}b = \mathbb{Z}c \quad \text{with } c = \operatorname{lcm}(a,b).$$

Proof: $x \in \mathbb{Z}a \cap \mathbb{Z}b \Leftrightarrow x$ is a multiple of both a and $b \Leftrightarrow x$ is a multiple of $c \Leftrightarrow x \in \mathbb{Z}c$.

Finally, the *product* of two ideals corresponds to the product of the generators:

$$\mathbb{Z}a \cdot \mathbb{Z}b = \mathbb{Z}ab.$$

The easy proof of this is left as an exercise. ∎

If R is a commutative ring with 1 then:

$$(a_1, \ldots, a_n) \cdot (b_1, \ldots b_m) = (a_1 b_1, \ldots, a_i b_j, \ldots, a_n b_m)$$

as is easily verified using the definitions. Moreover

$$(a, b) = (a + rb, b)$$

for all $a, b, r \in R$ (verify!), in this way one can 'eliminate' generators.

II.4.11 Example. In the ring $\mathbb{Z}[\sqrt{-5}]$ one finds:

$$\begin{aligned}
&(2, 1+\sqrt{-5}) \cdot (3, 1-\sqrt{-5}) = \\
&= (6, 2-2\sqrt{-5}, 3+3\sqrt{-5}, 6) \\
&= (6, 2-2\sqrt{-5}, 1+5\sqrt{-5}) \\
&= (6, 6\sqrt{-5}, 2-2\sqrt{-5}, 1+5\sqrt{-5}) \\
&= (6, 2-2\sqrt{-5}, 1-\sqrt{-5}) \\
&= ((1+\sqrt{-5}) \cdot (1-\sqrt{-5}), 2(1-\sqrt{-5}), 1-\sqrt{-5}) \\
&= (1-\sqrt{-5}).
\end{aligned}$$

Neither of the ideals $(2, 1+\sqrt{-5})$ and $(3, 1-\sqrt{-5})$ is principal (see Exercise 17 on page 32), however the calculation above shows that their product *is* a principal ideal. Multiplying ideals plays an important role in (algebraic) number theory. ∎

II.4.12 Theorem. (Chinese remainder theorem for rings). *Let R be a commutative ring with 1 and suppose I, J are coprime ideals of R, so $I+J=R$.*
Then $I\cap J = I\cdot J$ and there is an isomorphism of rings

$$R/(I\cdot J)\cong (R/I)\times(R/J),\qquad a+(I\cdot J)\mapsto (a+I,a+J).$$

Proof. We first show that $I+J=R$ implies $I\cap J = I\cdot J$.

The inclusion $I\cap J\supset I\cdot J$ is true in general. For the other inclusion, by $I+J=R$ there exist $x\in I$, $y\in J$ with $x+y=1$. Hence every $z\in I\cap J$ satisfies

$$z = z\cdot 1 = z\cdot(x+y) = x\cdot z + z\cdot y$$

where $x\cdot z\in I\cdot J$ (since $x\in I, z\in J$) and $z\cdot y\in I\cdot J$ (since $z\in I, y\in J$). Therefore $z\in I\cdot J$. This proves $I\cap J = I\cdot J$.

Take $\phi_1 : R\to R/I$ and $\phi_2 : R\to R/J$ the canonical ring homomorphisms with kernel I and J, respectively. Define

$$\phi : R\to (R/I)\times(R/J)\qquad \phi(a) := (\phi_1(a),\phi_2(a)).$$

We claim that ϕ is a surjective ring homomorphism with kernel $I\cdot J$ is. Using this, the desired isomorphism $R/I\cdot J\cong (R/I)\times(R/J)$ follows immediately from the first isomorphism theorem II.3.7.

(i) ϕ is a ring homomorphism:

$$\phi(ab) = (\phi_1(ab),\phi_2(ab)) = (\phi_1(a)\phi_1(b),\phi_2(a)\phi_2(b))$$

(since ϕ_1,ϕ_2 are ring homomorphisms). Here the right hand side equals

$$\begin{aligned} &= (\phi_1(a),\phi_2(a))\cdot(\phi_1(b),\phi_2(b))\\ &= \phi(a)\phi(b)\end{aligned}$$

by definition of the multiplication on a product of two rings. So $\phi(ab)=\phi(a)\phi(b)$. In a similar way one verifies $\phi(a+b)=\phi(a)+\phi(b)$.

(ii) $\mathrm{Ker}(\phi) = I\cdot J$: we have $a\in ker(\phi)$ precisely when $(\phi_1(a),\phi_2(a)) = (0,0)$, which is equivalent to $a\in \mathrm{Ker}(\phi_1)=I \wedge a\in \mathrm{Ker}(\phi_2)=J$, hence to $a\in I\cap J = I\cdot J$.

(iii) ϕ is surjective: take $x+y=1$ as above with $x\in I, y\in J$. Then

$$\phi_1(x)=0,\quad \phi_2(y)=0$$

and (using $x=1-y$)

$$\begin{aligned}&\phi_2(x) = \phi_2(1)-\phi_2(y) = 1-0 = 1\in R/J,\\ &\phi_1(y) = \phi_1(1-x) = 1\in R/I.\end{aligned}$$

This shows $\phi(x)=(0,1),\quad \phi(y)=(1,0)$. Now let ($a$ mod I, b mod J (with $a,b\in R$) be an arbitrary element of $(R/I)\times(R/J)$ and put $c=bx+ay$. Then

$$\begin{aligned}\phi_1(c) &= \phi_1(b)\phi_1(x)+\phi_1(a)\phi_1(y)\\ &= \phi_1(b)\cdot 0+\phi_1(a)\cdot 1\\ &= \phi_1(a)\end{aligned}$$

and similarly $\phi_2(c)=\phi_2(b)$. Therefore $\phi(c)=(a \bmod I, b \bmod J)$ which shows that ϕ is surjective.

This proves the Chinese remainder theorem. ∎

II.4.13 Corollary. *Let $n,m\in\mathbb{Z}$ be relative prime. There is a ring isomorphism*

$$\mathbb{Z}/nm\mathbb{Z}\cong(\mathbb{Z}/n\mathbb{Z})\times(\mathbb{Z}/m\mathbb{Z}),\qquad a+nm\mathbb{Z}\mapsto(a+n\mathbb{Z},a+m\mathbb{Z}).$$

Proof. This follows from II.4.12 using $\mathbb{Z}n + \mathbb{Z}m = \mathbb{Z}\gcd(n,m) = \mathbb{Z}$ (Example II.4.10). ∎

Note that the condition that n, m are coprime is necessary. For example we have $\mathbb{Z}/4\mathbb{Z} \ncong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ since the additive groups of these rings are not isomorphic.

II.4.14 Example. Let $R = \mathbb{Q}[X]$ and take

$$I = \mathbb{Q}[X] \cdot (X-1) \qquad \text{en} \qquad J = \mathbb{Q}[X] \cdot (X+1).$$

Then

$$-\frac{1}{2}(X-1) \in I, \quad \frac{1}{2}(X+1) \in J, \qquad \text{and} \quad -\frac{1}{2}(X-1) + \frac{1}{2}(X+1) = 1,$$

so the ideals I and J are coprime. We have

$$I \cdot J = \mathbb{Q}[X] \cdot (X+1)(X-1) = \mathbb{Q}[X] \cdot (X^2 - 1),$$

and II.4.12 implies

$$\mathbb{Q}[X]/\mathbb{Q}[X](X^2-1) \cong (\mathbb{Q}[X]/I) \times (\mathbb{Q}[X]/J).$$

Furthermore II.2.10 shows $\mathbb{Q}[X]/I \cong \mathbb{Q}$, $f \mapsto f(1)$ and $\mathbb{Q}[X]/J \cong \mathbb{Q}$, $f \mapsto f(-1)$, hence

$$\mathbb{Q}[X]/\mathbb{Q}[X](X^2-1) \cong \mathbb{Q} \times \mathbb{Q}, \qquad f + (X^2-1) \mapsto (f(1), f(-1)).$$

▬

II.4.15 Example. If $R = R_1 \times R_2$ where R_1, R_2 are rings with 1, then (1,0) and (0,1) are idempotents of R. We will now show that in case R is commutative, all idempotents are obtained from writing R as a product of two rings.

So let R be a commutative ring with 1 and $e \in R$ an idempotent. We apply II.4.12 to

$$I = R \cdot e, \quad J = R \cdot (1-e).$$

From $e + (1-e) = 1$ it follows that I and J are coprime. We have $I \cdot J = R(e - e^2) = \{0\}$, since e is idempotent. Hence $R/I \cdot J \cong R/\{0\} \cong R$ and II.4.12 yields

$$R \cong (R/Re) \times (R/R(1-e)),$$

with an isomorphism sending e to $(0,1)$ and $1-e$ to $(1,0)$. Apparently $1-e$ is an idempotent as well, as one may also verify directly.

We conclude that for any commutative ring R with 1 we have a $1-1$ correspondence between pairs of idempotents $\{e, 1-e\}$ of R and ways to write R as a product of two rings R_1 and R_2.

An explicit example: using $R = \mathbb{Z}/6\mathbb{Z}$, $e = \bar{4}$, one obtains $\mathbb{Z}/6\mathbb{Z} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ (compare II.4.13). ▬

II.5 Exercises

1. Let R be a unitary ring. Show that a unique unitary ring homomorphism $f : \mathbb{Z} \to R$ exists.
   The non-negative generator of $\mathrm{Ker}(f)$ *is called the* characteristic *of* R*, notation:* $\mathrm{char}(R)$.
2. Prove that the characteristic of a domain is either 0 or a prime number.
3. Show that the following rings have no ring automorphism except the identity:
$$\mathbb{Z}, \quad \mathbb{Z}/n\mathbb{Z} \,(\text{for } n \in \mathbb{Z}_{>0}), \quad \mathbb{Q}.$$
4. Let σ be a ring automorphism of $\mathbb{R}$.
   (a) Prove: $x > 0 \Rightarrow \sigma(x) > 0$.
   (b) Prove: $\sigma = id_{\mathbb{R}}$.
5. Show that $\mathbb{C}$ admits a ring automorphism different from the identity map.
6. Let K be a field and $R \subset K$ a subring with $1 \in R$. Assume that every element of K can be written as as^{-1} with $a, s \in R, s \neq 0$. Show: K is isomorphic to the field of fractions $Q(R)$ of R.
7. Is $\det : M(n, \mathbb{R}) \to \mathbb{R}$ a ring homomorphism?
8. Let $f : R_1 \to R_2$ be a unitary ring homomorphism. Show that $g = f|R_1^\times$ is a group homomorphism $R_1^\times \to R_2^\times$, and show by means of an example that g need not be surjective in case f is.
9. Let $G = \{1, \sigma\}$ be a multiplicatively written group with two elements. Define $f : \mathbb{R}[G] \to \mathbb{R} \times \mathbb{R}$ by $f(a + b\sigma) = (a + b, a - b)$, for $a, b \in \mathbb{R}$. Show that f is a ring isomorphism.
10. (a) Let R be a domain and $R' \subset R$ a subring with $1 \in R'$. Show that $Q(R')$ can be regarded as a subring of $Q(R)$ (hint: verify that $\frac{a}{b} \mapsto \frac{a}{b}$, with the first "fraction" in $Q(R')$ and the second one in $Q(R)$, is a well defined injective ring homomorphism).
    (b) Prove for any domain R that
$$R \to Q(R) : r \mapsto \frac{r}{1} \text{ is surjective} \iff R \text{ is a field.}$$
    (c) Suppose $m \in \mathbb{Z}$ satisfies $\sqrt{m} \notin \mathbb{Z}$. Show that $Q(\mathbb{Z}[\sqrt{m}])$ can be identified with $\mathbb{Q}[\sqrt{m}]$.
11. Prove: $\mathrm{End}(\mathbb{Z}^+) \cong \mathbb{Z}$, $\mathrm{End}(\mathbb{Q}^+) \cong \mathbb{Q}$, $\mathrm{End}((\mathbb{Z}/n\mathbb{Z})^+) \cong \mathbb{Z}/n\mathbb{Z}$ as rings.
12. Let $(A, +, 0)$ be an abelian group and $B = \{a \in A : a \text{ has finite order}\}$. Define $I \subset \mathrm{End}(A)$ by
$$I = \{\sigma \in \mathrm{End}(A) : \sigma(x) = 0 \text{ for all } x \in B\}.$$
    Prove that I is an ideal of $\mathrm{End}(A)$ and $\mathrm{End}(A)/I$ is isomorphic to a subring of $\mathrm{End}(B)$.
13. Let R be a ring. For $a \in R$ we define $\lambda_a, \rho_a : R \to R$ by $\lambda_a(x) = ax, \rho_a(x) = xa$.
    (a) Show that $\lambda_a, \rho_a \in \mathrm{End}(R^+)$ for all $a \in R$.
    (b) Show that the map $f : R \to \mathrm{End}(R^+)$, $f(a) = \lambda_a$ is a ring homomorphism. Moreover prove that in case R is a ring with 1, then f is unitary and injective.
    (c) Show that the map $g : R^0 \to \mathrm{End}(R^+)$, $g(a) = \rho_a$ is a ring homomorphism, where R^0 denotes the 'opposite ring' defined in Exercise 11 on page 14.

14. A *Cauchy sequence* over $\mathbb{Q}$ is a sequence $(a_n)_{n=1}^{\infty}$, with $a_n \in \mathbb{Q}$ satisfying:

$$\forall \epsilon \in \mathbb{Q}_{>0} : \exists n_0 : \forall n, m > n_0 : \quad |a_n - a_m| < \epsilon.$$

The set of Cauchy sequences over $\mathbb{Q}$ forms a ring R, with component wise operations. A *zero sequence* is a sequence $(a_n)_{n=1}^{\infty}$ with $a_n \in \mathbb{Q}$ satisfying $\lim_{n\to\infty} a_n = 0$. Show that $I \subset R$ and that in fact I is an *ideal* of R. Moreover, show that $R/I \cong \mathbb{R}$.

15. Let R be a ring with 1 and G a group. Define $f : R[G] \to R$ by $f(\sum_{g\in G} a_g g) = \sum_{g\in G} a_g$. Show that f is a surjective ring homomorphism, and that the kernel $\mathrm{Ker}(f)$ is generated by $\{g - 1 : \ g \in G\}$.
16. Let $R = \mathbb{Z}[X]$ and put

$$\phi : \mathbb{Z}[X] \to \mathbb{Z}/2\mathbb{Z}, \qquad f \mapsto f(0) + 2\mathbb{Z}.$$

(a) Show that ϕ is a surjective ring homomorphism and that $\mathrm{Ker}(\phi) = (2, X)$.

(b) Prove that $(2, X)$ is not a principal ideal.

17. Let $R = \mathbb{Z}[\sqrt{-5}]$ and

$$\phi : \mathbb{Z}[\sqrt{-5}] \to \mathbb{Z}/3\mathbb{Z}, \qquad a + b\sqrt{-5} \mapsto \overline{a + b} \qquad (a, b \in \mathbb{Z}).$$

(a) Prove that ϕ is a surjective ring homomorphism.

(b) Prove that $\mathrm{Ker}(\phi) = (3, 1 - \sqrt{-5})$.

(c) Prove that $\mathrm{Ker}(\phi)$ is not a principal ideal, as follows: if $\mathrm{Ker}(\phi) = (x)$, then $3 = xy$ and $1 - \sqrt{-5} = xz$. Consider $N(xy)$ and $N(xz)$ with N as in I.2.5.)

(d) Prove similarly that $(2, 1 + \sqrt{-5})$ is not principal.

(e) Is the ideal $(3, 1 - \sqrt{-5}) \cdot (3, 1 - \sqrt{-5})$ principal?

18. Define $\varphi : \mathbb{Z}[i] \to \mathbb{F}_{13}$ by $\varphi(a + bi) = a + 5b \bmod 13$.
Show that φ is a ring homomorphism with $\mathrm{Ker}(\varphi)$ generated by 13 and $i - 5$. Find one generator for $\mathrm{Ker}(\varphi)$.
19. Let R be a ring and $I \subset R$ a left ideal containing a right unit. Show: $I = R$.
20. Let R_1 and R_2 be rings and $I = \{0\} \times R_2 \subset R_1 \times R_2$.

(a) Show that I is an ideal of $R_1 \times R_2$.

(b) Prove that $(R_1 \times R_2)/I \cong R_1$.

(c) In case R_2 is unitary, show that I is a principal ideal.

21. Let R_1 and R_2 be rings. Show that all ideals of $R_1 \times R_2$ are of the form $I_1 \times I_2$ with I_i an ideal of R_i $(i = 1, 2)$.
22. Let R be a ring with 1 having the property that $f : R \to R$, $f(x) = x^2$ is a ring homomorphism. Prove that R is commutative, and $\mathrm{char}(R) = 1$ of 2 (see Exercise II.5.1). Moreover, show that $\forall x \in \mathrm{Ker}(f) : \ 1 + x \in R^{\times}$.
23. (a) Show that

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}(2, \mathbb{R}) : \ b = d = 0 \right\}$$

is a left ideal but not a right ideal of $\mathrm{M}(2, \mathbb{R})$.

(b) Find a right ideal of $\mathrm{M}(2, \mathbb{R})$ that is not a left ideal.

24. Let $n \in \mathbb{Z}_{>0}$. In this exercise elements of $\mathrm{M}(n, \mathbb{R})$ are considered as $\mathbb{R}$-linear endomorphisms of $\mathbb{R}^n$. By W we denote an $\mathbb{R}$-linear subspace of $\mathbb{R}^n$.

(a) Show that $\{A \in \mathrm{M}(n, \mathbb{R}) : \forall w \in W : Aw = 0\}$ is a left ideal of $\mathrm{M}(n, \mathbb{R})$.

(b) Show that $\{A \in \mathrm{M}(n, \mathbb{R}) : \forall v \in \mathbb{R}^n : Av \in W\}$ is a right ideal of $\mathrm{M}(n, \mathbb{R})$.

(c) Prove: every left ideal of $\mathrm{M}(n, \mathbb{R})$ is of the form described in (a), and every right ideal of $\mathrm{M}(n, \mathbb{R})$ is of the form described in (b).

(d) Prove that $\{0\}$ and $\mathrm{M}(n,\mathbb{R})$ are the only two sided ideals of $\mathrm{M}(n,\mathbb{R})$.

25. Let $I \subset R$ be an ideal in a ring and write $\phi : R \to R/I$ for the corresponding natural map.

(a) Suppose $J' \subset R/I$ is an ideal. Show that

$$\phi^{-1}(J') := \{x \in R : \phi(x) \in J'\}$$

is an ideal of R. Note that $I \subset \phi^{-1}(J')$.

(b) Show that $J' \mapsto \phi^{-1}(J')$ defines a bijection between the ideals J' of R/I and the ideals J of R with $I \subset J$.

(c) Prove that any ideal J of R with $I \subset J$ satisfies $(R/I)/\phi(J) \cong R/J$.

26. Let K be a field. The *ring of dual numbers* over K, notation: $K[\epsilon]$, consists of the expressions $a + b\epsilon$ for $a, b \in K$, and addition and multiplication as follows:

$$(a + b\epsilon) + (c + d\epsilon) = (a + c) + (b + d)\epsilon,$$

$$(a + b\epsilon) \cdot (c + d\epsilon) = (ac) + (ad + bc)\epsilon$$

(in particular $\epsilon^2 = 0$), with $a, b, c, d \in K$.

(a) Prove: $K[\epsilon] \cong K[X]/(X^2)$.

(b) Prove that $K[\epsilon]$ contains precisely *three* ideals.

(c) Prove: $K[\epsilon]^\times \cong K^\times \times K^+$ (as groups).

27. Let R be a ring with $1 \neq 0$ and $I = R - R^\times$. Suppose that $\forall x \in I : \exists n \in \mathbb{Z}_{>0} : x^n = 0$. Show that I is an ideal of R, and that R/I is a division ring.

28. Let R be a ring and I an ideal of R. Suppose $R' \subset R$ is a subring. Prove:

(a) $R' \cap I$ is an ideal of R';

(b) $R' + I = \{r + s : r \in R', s \in I\}$ is a subring of R;

(c) I is an ideal of $R' + I$;

(d) $R'/(R' \cap I) \cong (R' + I)/I$.

29. Let R be a ring with 1. Define

$$[R,R] = \left\{ \sum_{i=1}^{n} r_i(x_i y_i - y_i x_i) : n \in \mathbb{Z}_{>0},\ r_i, x_i, y_i \in R \right\}.$$

Prove that $[R,R]$ is an ideal of R and $R/[R,R]$ is a commutative ring.

30. Let

$$R = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}(2,\mathbb{R}) : c = 0 \right\}$$

en

$$I = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \in \mathrm{M}(2,\mathbb{R}) : b \in \mathbb{R} \right\}.$$

Prove the following assertions:

(a) R is a subring of $\mathrm{M}(2,\mathbb{R})$;

(b) I is an ideal of R, and $R/I \cong \mathbb{R} \times \mathbb{R}$;

(c) R is not commutative but R/I is.

31. In $R = \mathbb{Z}[X]$ take $I = (2, X)$. Show that $X^2 + 4 \in I \cdot I$, and $X^2 + 4$ can not be written as xy with $x, y \in I$. Conclude that $\{xy : x, y \in I\}$ is not an ideal of R.

32. Let R be a ring and I, J ideals of R. Prove $(I + J) \cdot (I \cap J) \subset (I \cdot J) + (J \cdot I)$. Show that equality holds in case $R = \mathbb{Z}$.

33. Prove that II.4.12 also holds for non-commutative rings, provided one replaces in two spots $I \cdot J$ by $I \cdot J + J \cdot I$.

34. Let R be a ring with 1 and I_1, I_2, I_3 ideals of R.
Show: $I_1 + I_3 = R \wedge I_2 + I_3 = R \Rightarrow (I_1 \cdot I_2) + I_3 = R$.
35. (Chinese remainder theorem for more ideals). Suppose R is a commutative ring with 1, and $I_1, I_2, ..., I_t$ are ideals of R which are pairwise coprime, i.e., $I_i + I_j = R$ for $1 \leq i < j \leq t$. Prove: $R/(\prod_{i=1}^{t} I_i) \cong \prod_{i=1}^{t}(R/I_i)$. (Hint: show $(I_1 \cdot I_2 \cdot ... \cdot I_{t-1}) + I_t = R$ as in Exercise 34 above, and use mathematical induction with respect to t.)
36. Let R be a ring with 1 such that $1 + 1 \in R^\times$. Prove: $R[X]/R[X](X^2 - 1) \cong R \times R$.
37. Put $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a \equiv b \bmod 2\}$.
    (a) Verify that R is a subring of $\mathbb{Z} \times \mathbb{Z}$.
    (b) Show that $\mathbb{Z}[X]/\mathbb{Z}[X] \cdot (X^2 - 1) \cong R$.
    (c) Prove that $\mathbb{Z}[X]/\mathbb{Z}[X] \cdot (X^2 - 1)$ is *not* isomorphic to $\mathbb{Z} \times \mathbb{Z}$ (for example: find the idempotents in both rings).
    (d) Show that no $f \in \mathbb{Z}[X]$ exists with $f(1) = 1, f(-1) = 0$. (Do you see the relation between this and the other parts of this exercise?)
38. Let R be a commutative ring with 1, and $w_1, w_2, ..., w_m$ elements of R such that $w_i - w_j \in R^\times$ for all i, j with $1 \leq i < j \leq m$. Take $f = \prod_{i=1}^{m}(X - w_i) \in R[X]$. Prove: $R[X]/R[X]f \cong R \times R \times ... \times R$ (product of m copies of R).
39. Prove $\mathbb{Q}[X]/\mathbb{Q}[X](X^3 + X) \cong \mathbb{Q} \times \mathbb{Q}[X]/(X^2 + 1)$ and $\mathbb{R}[X]/\mathbb{R}[X](X^4 - 1) \cong \mathbb{R} \times \mathbb{R} \times \mathbb{C}$.
40. Suppose R is a commutative ring with 1. By $\mathrm{Id}(R)$ we will denote the set of all idempotents of R (including the trivial idempotents 0 and 1). Show that if $e_1, e_2 \in \mathrm{Id}(R)$ then also $e_1 + e_2 - 2e_1e_2 \in \mathrm{Id}(R)$ and $e_1e_2 \in \mathrm{Id}(R)$.
Show that $\mathrm{Id}(R)$ becomes a commutative ring with addition $\oplus$ and multiplication $\circ$ defined by $e_1 \oplus e_2 = e_1 + e_2 - 2e_1e_2$ and $e_1 \circ e_2 = e_1e_2$.
Under which conditions is $\mathrm{Id}(R)$ a subring of R?

III RINGS OF POLYNOMIALS

III.1 Polynomials

Let R be a ring. We will define a ring $R[X]$ called the polynomial ring over R in the variable X. A *polynomial* with coefficients in R is a formal expression

$$\sum_{i=0}^{<\infty} a_i X^i = a_0 + a_1 X + a_2 X^2 + \ldots, \qquad a_i \in R.$$

Here the a_i are called the *coefficients* of the polynomial $\sum_{i=0}^{<\infty} a_i X^i$. Two polynomials $\sum_{i=0}^{N} a_i X^i$ and $\sum_{i=0}^{N} b_i X^i$ are by definition *equal* if and only if $\forall i \geq 0 : \ a_i = b_i$. Moreover with $0 \leq N < M$ we say that $\sum_{i=0}^{N} a_i X^i = \sum_{i=0}^{M} a_i X^j$ in case $a_j = 0$ for all j with $N < j \leq M$. Instead of the symbol/variable X one also uses other symbols such as $Y, Z, U, T, X_0, X_1, \ldots$. Terms $a_i X^i$ with $a_i = 0$ are usually deleted. Instead of $1 \cdot X^i$ one writes X^i, and $(-a) \cdot X^i$ is written $-aX^i$. For example:

$$1 - 2X + X^3 = 1 + (-2) \cdot X + 0 \cdot X^2 + 1 \cdot X^3.$$

One often denotes a polynomial $\sum a_i X^i$ by a letter such as f, or by $f(X)$ if it should be made clear which variable is used.

III.1.1 Remark. The description of polynomials as formal finite expressions is by far the most intuitive one. However, alternatively polynomials may also be introduced as follows.

Let $\mathbb{N} = \mathbb{Z}_{\geq 0}$ and let R be a ring. A '*polynomial*' is a function

$$a \quad : \quad \mathbb{N} \longrightarrow R$$

such that some N exists with $a(n) = 0$ for all $n > N$. Identifying a polynomial $\sum_{i=0}^{N} a_i X^i$ with the function ('*polynomial*') $a : \mathbb{N} \to R$ given by $a(n) = a_n$ for $n \leq N$ and $a(n) = 0$ when $n > N$, it is easy to change the one definition into the other.

The *degree* $\deg(f)$ of a polynomial $f = \sum_{i=0}^{N} a_i X^i$ is the maximal n with $a_n \neq 0$; so $\deg(1 - 2X + X^3) = 3$. For the *zero polynomial* $0 = \sum_{i=0}^{N} 0 \cdot X^i$ one defines $\deg(0) = -\infty$ (some texts use alternative definitions for $\deg(0)$, however).

The *j-th coefficient* of a polynomial $f = \sum_{i=0}^{N} a_i X^i$ is a_j. The *constant* coefficient is the coefficient a_0. A *constant polynomial* f is a polynomial with $\deg(f) \leq 0$, i.e., $a_n = 0$ for $n \geq 1$. If $f \neq 0$ and $n = \deg(f)$, then a_n is called the *leading coefficient* of f. A polynomial with leading coefficient 1 is called a *monic* polynomial.

We now define operations addition and multiplication on the set of polynomials with coefficients in a ring R. The *sum* of two polynomials of degree $\leq N$ is defined as

$$\left(\sum_{i=0}^{N} a_i \cdot X^i\right) + \left(\sum_{i=0}^{N} b_i \cdot X^i\right) = \sum_{i=0}^{\infty} (a_i + b_i) \cdot X^i.$$

Note that considered as functions $\mathbb{N} \to R$ this is simply the pointwise addition: $a + b = c$ with $c(n) = a(n) + b(n)$ for all $n \in \mathbb{N}$.

The *multiplication* of polynomials is determined by the rule

$$(a_i X^i) \cdot (b_j X^j) = (a_i \cdot b_j) X^{i+j}$$

and the fact that one wants the distributive law to hold; this leads to the definition

$$\left(\sum_{i=0}^{N} a_i \cdot X^i\right) \cdot \left(\sum_{j=0}^{M} b_j \cdot X^j\right) = \sum_{k=0}^{N+M} \left(\sum_{i+j=k} a_i b_j\right) \cdot X^k,$$

using the convention that $a_i = 0$ if $i > N$ and similarly $b_j = 0$ if $j > M$.

III.1.2 Example.

$$\begin{aligned}
&(7 + 3X)(5 - X + 2X^2) = \\
&= 7 \cdot 5 + (7 \cdot -1 + 3 \cdot 5)X + (7 \cdot 2 + 3 \cdot -1)X^2 + 3 \cdot 2X^3 \\
&= 35 + 8X + 11X^2 + 6X^3.
\end{aligned}$$

—■

III.1.3 Remark. Considered as functions $a, b : \mathbb{N} \to R$ the multiplication of '*polynomials*' as defined here, is usually called the *convolution* of the functions a and b, denoted as $a * b$. So by definition $a * b : \mathbb{N} \to R$ is the function defined by

$$(a * b)(n) = \sum_{i=0}^{n} a(i)b(n-i) = a(0)b(n) + a(1)b(n-1) + \ldots + a(n)b(0).$$

The set of all polynomials with coefficients in R is denoted by $R[X]$. Note that if $f, g \in R[X]$ then

$$\deg(f \cdot g) \leq \deg(f) + \deg(g).$$

Equality does not hold in general: for example take $f = \overline{2} \cdot X^2 + \overline{1} \in (\mathbb{Z}/4\mathbb{Z})[X]$. Then $\deg(f) = 2$ and $f^2 = (\overline{2} \cdot X^2 + \overline{1}) \cdot (\overline{2} \cdot X^2 + \overline{1}) = \overline{1}$. Hence $\deg(f^2) = 0 < 4 = \deg(f) + \deg(f)$ in this example.

III.1.4 Theorem. *The set $R[X]$ with addition and multiplication defined above defines a ring called the polynomial ring in one variable over R.*

Using the injective ring homomorphism

$$R \hookrightarrow R[X], \qquad r \mapsto r + 0 \cdot X + \ldots + 0 \cdot X^i + \ldots,$$

one considers R as a subring of $R[X]$.

Proof. The proof is a straightforward verification of the axioms. As an illustration of this we check (R3), the associativity of the multiplication:

$$\begin{aligned}
\left(\left(\textstyle\sum_{i=0}^{N} a_i \cdot X^i\right)\left(\textstyle\sum_{j=0}^{M} b_j \cdot X^j\right)\right)\left(\textstyle\sum_{k=0}^{L} c_k \cdot X^k\right) &= \\
\left(\textstyle\sum_{l=0}^{N+M} \left(\textstyle\sum_{i+j=l} a_i b_j\right) \cdot X^l\right)\left(\textstyle\sum_{k=0}^{L} c_k \cdot X^k\right) &= \\
\textstyle\sum_{m=0}^{N+M+L} \left(\textstyle\sum_{k+l=m} \left(\textstyle\sum_{i+j=l} a_i b_j\right) c_k\right) \cdot X^m &= \\
\textstyle\sum_{m=0}^{N+M+L} \left(\textstyle\sum_{i+j+k=m} a_i b_j c_k\right) \cdot X^m,&
\end{aligned}$$

and analogously one shows

$$\left(\sum_{i=0}^{N} a_i \cdot X^i\right)\left((\sum_{j=0}^{M} b_j \cdot X^j)(\sum_{k=0}^{L} c_k \cdot X^k)\right) = \sum_{m=0}^{N+M+L}\left(\sum_{i+j+k=m} a_i b_j c_k\right)\cdot X^m.$$

This proves (R3). We leave (R1), (R2), and (R4) to the reader.

The remaining assertion in the theorem is also easy to check. ∎

In case R is commutative, so is $R[X]$. If R is a ring with 1 then this 1 is the unit element of $R[X]$ as well. If R has no zero divisors, then the same holds for $R[X]$ (see Exercise 1 on page 49) and

$$\deg(f \cdot g) = \deg(f) + \deg(g) \qquad \text{for} \quad f, g \in R[X],$$

where we use the convention $-\infty + n = n + (-\infty) = -\infty + (-\infty) = -\infty$. In particular if R is a domain (Definition I.2.13) then $R[X]$ is a domain as well.

III.1.5 Notation. Inductively one defines the polynomial ring in $n \geq 1$ variables over R as

$$R[X_1, X_2, \ldots, X_n] := (R[X_1, \ldots, X_{n-1}])[X_n].$$

Elements of $R[X_1, X_2, \ldots, X_n]$ are therefore expressions

$$f = g_0 + g_1 X_n + g_2 X_n^2 + \ldots = g_N X^N, \qquad g_i \in R[X_1 \ldots, X_{n-1}].$$

One can also write f as a finite sum

$$f = \sum_{i_1 \geq 0, i_2 \geq 0, \ldots, i_n \geq 0} a_{i_1 i_2 \ldots i_n} X_1^{i_1} X_2^{i_2} \ldots X_n^{i_n}$$

with coefficients $a_{i_1 i_2 \ldots i_n} \in R$. Sometimes the "multi-index"-notation

$$f = \sum_I a_I X^I$$

where the 'multi-index' $I = (i_1, i_2, \ldots, i_n)$ runs over a finite subset of $(\mathbb{Z}_{\geq 0})^n$ and X^I is an abbreviation of $X_1^{i_1} X_2^{i_2} \ldots X_n^{i_n}$.

For polynomials in n variables one defines various notions of 'degree'. Given any j with $1 \leq j \leq n$ the *degree in* X_j of

$$f = \sum_{i_1 \geq 0, i_2 \geq 0, \ldots, i_n \geq 0} a_{i_1 i_2 \ldots i_n} X_1^{i_1} X_2^{i_2} \ldots X_n^{i_n} \neq 0$$

is defined by

$$\deg_j(f) := \max\{m \in \mathbb{Z}_{\geq 0} : \exists i_1, \ldots, i_n : a_{i_1 \ldots i_n} \neq 0 \text{ and } i_j = m\}$$

(so $\deg_j$ is the 'highest power' of X_j 'really appearing' in f). The *total degree* of $f \neq 0$ is defined by

$$\operatorname{tdeg}(f) = \max\left\{m \in \mathbb{Z}_{\geq 0} : \exists i_1, \ldots i_n : a_{i_1 \ldots i_n} \neq 0 \text{ and } \sum_{j=1}^{n} i_j = m\right\}.$$

Finally, in the special case $f = 0$ we put $\deg_j(0) = \operatorname{tdeg}(0) = -\infty$.

III.1.6 Example. $f = X_1 X_2^4 - X_1^2 X_2^2 \in \mathbb{Z}[X_1, X_2]$ satisfies $\deg_1(f) = 2, \deg_2(f) = 4$, and $\operatorname{tdeg}(f) = 5$. ∎

III.2 Evaluation homomorphisms

It is well known that a polynomial $f \in \mathbb{R}[X]$ can be evaluated in any real and even in any complex number z. For fixed $z \in \mathbb{C}$ one obtains in this way a map $\mathbb{R}[X] \to \mathbb{C}$ given by $f \mapsto f(z)$. Similarly given $f \in \mathbb{Z}[X]$ one may take $a \in \mathbb{Z}$, evaluate f in a and subsequently consider $\overline{f(a)} \in \mathbb{Z}/n\mathbb{Z}$. In this way one obtains a map $\mathbb{Z}[X] \to \mathbb{Z}/n\mathbb{Z}$ given by $f \mapsto \overline{f(a)}$. As is not hard to verify, the same map is obtained by first reducing the coefficients of f modulo n and subsequently evaluating the resulting polynomials in $(\mathbb{Z}/n\mathbb{Z})[X]$ at $\overline{a}$. The next result generalises these examples.

III.2.1 Theorem. (The evaluation homomorphism) *Let R and S be rings.*

(a) Any ring homomorphism $\phi : R \longrightarrow S$ induces a ring homomorphism

$$\Phi : R[X] \to S[X], \qquad \Phi : a_0 + \ldots + a_n X^n \mapsto \phi(a_0) + \ldots + \phi(a_n) X^n.$$

(b) If $s \in S$ satisfies $st = ts$ for all $t \in S$, then the map

$$S[X] \to S, \quad a_0 + a_1 X + \ldots + a_n X^n \mapsto a_0 + a_1 s + \ldots + a_n s^n$$

is a ring homomorphism. We usually write this as $f \mapsto f(s)$.

(c) If $s \in S$ satisfies $s\phi(r) = \phi(r)s$ for all $r \in R$ then the composition

$$\mathrm{ev}_s : R[X] \xrightarrow{\Phi} S[X] \overset{f \mapsto f(s)}{\longrightarrow} S$$

is a ring homomorphism.

III.2.2 Notation. The ring homomorphism ev_s in Theorem III.2.1 is called the *evaluation homomorphism* in s. Note that it depends on the choice of the ring homomorphism $R \to S$ although this is not reflected in the notation.

Proof. (a): given $f = a_0 + a_1 X + \ldots$, $g = b_0 + b_1 X + \ldots \in R[X]$ we have to show

$$\Phi(f+g) = \Phi(f) + \Phi(g) \quad \text{and} \quad \Phi(fg) = \Phi(f)\Phi(g).$$

As $f + g = (a_0 + b_0) + (a_1 + b_1)X + (a_2 + b_2)X^2 + \ldots$ we have

$$\begin{aligned} \Phi(f+g) &= \phi(a_0 + b_0) + \phi(a_1 + b_1)X + \phi(a_2 + b_2)X^2 + \ldots \\ &= \phi(a_0) + \phi(a_1)X + \phi(a_2)X^2 + \ldots + \phi(b_0) + \phi(b_1)X + \ldots \\ &= \Phi(f) + \Phi(g). \end{aligned}$$

For the second property of ring homomorphisms we write $fg = c_0 + c_1 X + c_2 X^2 + \ldots$, with $c_k = a_0 b_k + a_1 b_{k-1} + \ldots + a_k b_0$. Then $\phi(c_k) = \phi(a_0)\phi(b_k) + \ldots + \phi(a_k)\phi(b_0)$, which implies $\Phi(fg) \overset{\text{def}}{=} \phi(c_0) + \phi(c_1)X + \phi(c_2)X + \ldots = \Phi(f)\Phi(g)$. This shows (a).

To prove (b) we need to verify (using $st = ts$ for all $t \in S$) that

$$(f+g)(s) = f(s) + g(s), \qquad (fg)(s) = f(s) \cdot g(s) \qquad \forall f, g \in S[X].$$

The proof of the first condition is straightforward and hence we omit it. For the second condition note that $st = ts$ for all $t \in S$ implies in particular

$$(a_i s^i)(b_j s^j) = a_i b_j s^{i+j} \qquad \forall a_i, b_j \in S.$$

Writing $f = a_0 + a_1 X + \ldots$ and $g = b_0 + b_1 X + \ldots$ then yields

$$\begin{aligned} f(s) \cdot g(s) &= (a_0 + a_1 s + \ldots) \cdot (b_0 + b_1 s + \ldots) \\ &= a_0 b_0 + (a_0 b_1 + b_0 a_1)s + \ldots + (\textstyle\sum_{i+j=k} a_i b_j) s^k + \ldots \\ &= (fg)(s). \end{aligned}$$

Part (c) of the proof follows from the general fact that the composition of two ring homomorphisms is again a ring homomorphism. Here one should observe that $\mathrm{ev}_s : S[X] \to S$ is *not* necessarily a ring homomorphism (namely, s a priori only commutes with the elements in $\phi(R)$). However ev_s restricted to $\Phi(R[X])$ is a ring homomorphism. ∎

III.2.3 Remark. For $r := (r_1, \ldots, r_n) \in R^n$ we obtain an analogous evaluation homomorphism

$$\mathrm{ev}_r : R[X_1, \ldots, X_n] \longrightarrow R, \qquad \mathrm{ev}_r(f) := f(r_1, \ldots, r_n)$$

provided $r_i s = s r_i$ holds for all $s \in R$.

III.2.4 Example. Suppose K is a field, so in particular K is commutative. Let V be a vector space over K and write $\mathrm{End}_K(V)$ for the ring of K-linear maps from V to V. Fixing any $A \in \mathrm{End}_K(V)$ we will define an evaluation homomorphism $\mathrm{ev}_A : K[X] \to \mathrm{End}_K(V)$, which means we substitute A for the variable X in all polynomials $f \in K[X]$; the ring homomorphism property then says $f(A) \cdot g(A) = (fg)(A)$ and $f(A) + g(A) = (f+g)(A)$. We will also discuss the kernel of ev_A.

To define ev_A using Theorem III.2.1 we first need to define a ring homomorphism $\phi : K \to \mathrm{End}_K(V)$ such that $\phi(\lambda)A = A\phi(\lambda)$ for all $\lambda \in K$, $A \in \mathrm{End}_K(V)$. Put

$$\phi : K \longrightarrow \mathrm{End}_K(V), \qquad \phi : \lambda \mapsto \lambda I$$

with I the identity on V (so $(\lambda I)v := \lambda \cdot v$ for all $v \in V$). Then for all $A \in \mathrm{End}_K(V)$ and all $v \in V$ one finds

$$\big(\phi(\lambda)A\big)v = \phi(\lambda)(Av) = \lambda(Av) = A(\lambda v) = A\big(\phi(\lambda)v\big),$$

showing that indeed $\phi(\lambda)A = A\phi(\lambda)$ for all $\lambda \in K$, $A \in \mathrm{End}_K(V)$.

We write λA instead of $(\lambda I) \cdot A$. (In case $V = K^n$, note that $\mathrm{End}_K(V) = \mathrm{M}(n, K)$, the ring of $n \times n$ matrices with coefficients in K, and then λA is the matrix obtained by multiplying all coefficients of A by λ.)

Fix $A \in \mathrm{End}_K(V)$. Using Theorem III.2.1 one obtains the evaluation homomorphism in A:

$$\mathrm{ev}_A : K[X] \longrightarrow \mathrm{End}_K(V), \qquad \mathrm{ev}_A(f) := f(A),$$

and instead of $\mathrm{ev}_A(f) = \phi(a_0) + \phi(a_1)A + \phi(a_2)A^2 + \ldots$ we simply write

$$f(A) := a_0 I + a_1 A + a_2 A^2 + \ldots.$$

The image $\mathrm{ev}_A(K[X])$ we denote by $K[A]$. This is a subring of $\mathrm{End}_K(V)$. Since $K[X]$ is commutative, so is $K[A]$. In case the dimension of the vector space V is at least 2, the ring $\mathrm{End}_K(V)$ is not commutative. As a consequence, for $\dim_K(V) \geq 2$ the map ev_A is *not* surjective.

We now consider the special case $\dim_K(V) < \infty$. We will show that in this case ev_A is not injective. Write $\dim_K V = n$, then $\dim_K \mathrm{End}_K(V) = n^2$ (choose a basis for V over K, then $\mathrm{End}_K(V) \cong \mathrm{M}(n, K)$). The $n^2 + 1$ elements $I, A, A^2, \ldots, A^{n^2} \in \mathrm{End}_K(V)$ are therefore linearly dependent. Hence $c_i \in K$ exist, not all zero, such that

$$c_0 + c_1 A + c_2 A^2 + \ldots + c_{n^2} A^{n^2} = 0.$$

This shows $g(A) = 0$ with $g = c_0 + c_1 X + \ldots + c_{n^2} X^{n^2} \in K[X]$, so $g \in \mathrm{Ker}(\mathrm{ev}_A)$ and $g \neq 0$. As a consequence, ev_A is not injective.

You probably recall from a course in Linear Algebra that the eigenvalue polynomial (characteristic polynomial) P_A of A is in the kernel of ev_A,

$$P_A(X) := \det(A - X \cdot I) \quad \in K[X].$$

This result is called the Cayley-Hamilton theorem, named after the British mathematician Arthur Cayley (1821–1895) and the Irish mathematician William Rowan Hamilton (1805–1865). The degree of P_A is n. We will study the kernel of ev_A more extensively in Example III.4.4. ∎

III.2.5 Remark. Given a ring R, every polynomial $f = a_0 + a_1X + a_2X^2 + \ldots \in R[X]$ gives rise to a function, namely $\rho_f : R \to R$ given by

$$\rho_f : R \longrightarrow R, \qquad r \mapsto \rho_f(r) := a_0 + a_1r + a_2r^2 + \ldots.$$

It may happen that $f \neq g$ while $\rho_f(r) = \rho_g(r)$ for all $r \in R$. Hence the map

$$\rho : R[X] \longrightarrow \mathrm{Maps}(R,R), \qquad f \mapsto \rho_f,$$

is not necessarily injective.

As an example, consider $R = \mathbb{Z}/2\mathbb{Z}$. In this case the polynomials X and X^2 yield the same function on $\mathbb{Z}/2\mathbb{Z}$ (since $\bar{0} = \bar{0}^2$, $\bar{1} = \bar{1}^2$), however the polynomials X and X^2 are different (they even have different degree).

One should realise, as this example shows, that there is a difference between polynomials f and the functions ρ_f defined by them.

III.3 Division with remainder for polynomials

We now present a technique which, among various other applications, will greatly simplify the determination of the kernel of an evaluation homomorphism. This technique called *division with remainder* for polynomials is analogous to the division with remainder for integers.

III.3.1 Theorem. *Let R be a ring with 1 and $f, g \in R[X]$. Assume that $g \neq 0$ and that the leading coefficient of g is a unit of R.*

Then unique $q, r \in R[X]$ exist such that

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

One calls q the quotient and r the remainder when dividing by g.

III.3.2 Notation. The polynomials q and r in Theorem III.3.1 are called the *quotient* and the *remainder* of f upon division by g.

Proof. First we show the existence of q and r. Let $n = \deg(f)$ and $m = \deg(g)$. By assumption we have $m \geq 0$. We show existence, for fixed g, by induction w.r.t. n.

If $n < m$ one takes $q = 0$, $r = f$; this is the first step of the induction argument.

Now take $n \geq m$ and assume (induction hypothesis) that existence holds for all $f \in R[X]$ of degree $< n$. Take $f \in R[X]$ of degree n and let a be the leading coefficient of f, and b the leading coefficient of g. By assumption b is a unit, hence $c \in R$ exists with $cb = 1$. The polynomial $acX^{n-m} \cdot g$ than has degree n and leading coefficient $a \cdot cb = a$, equal to the leading coefficient of f. Hence

$$f_1 := f - acX^{n-m} \cdot g$$

has degree strictly *smaller* than n: the n-th degree terms in $f - acX^{n-m} \cdot g$ cancel. Applying the induction hypothesis to f_1 yields $q_1, r_1 \in R[X]$ such that

$$f_1 = q_1g + r_1, \qquad \deg(r_1) < \deg(g).$$

As a consequence

$$f = f_1 + acX^{n-m}g = \left(acX^{n-m} + q_1\right)\cdot g + r_1.$$

Hence $q := acX^{n-m} + q_1$ and $r := r_1$ satisfy

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

as desired.

It remains to show the unicity of q and r. Suppose that also $f = q'g + r'$ and $\deg(r') < \deg(g)$. Then

$$(q - q')g = r' - r.$$

Here the degree of the right hand side is less than $\deg(g)$. If $q \neq q'$, then the degree of the left hand side would be at least $\deg(g)$, since the leading coefficient of g is a unit. Hence

$$\deg(g) \leq \deg((q - q')g) = \deg(r' - r) < \deg(g).$$

This is a contradiction, and therefore $q = q'$ which implies $r' - r = 0$ so $r = r'$.

This proves Theorem III.3.1. Note that the the argument does not require the ring R to be commutative. ∎

III.3.3 Example. In practice dividing polynomials is done using '*long division*'. Take for instance $R = \mathbb{Z}$ and

$$f = X^4 - X^3 - 2X^2 + 3X - 4, \qquad g = X^2 - 1 \in \mathbb{Z}[X].$$

The quotient q and the remainder r are determined below:

$$\begin{array}{rcccccl}
X^2-1\,/ & X^4 & -X^3 & -2X^2 & +3X & -4 & \backslash\, X^2-X-1 \\
 & X^4 & & -X^2 & & & \\
\hline
 & & -X^3 & -X^2 & & & \\
 & & -X^3 & & +X & & \\
\hline
 & & & -X^2 & +2X & & \\
 & & & -X^2 & & +1 & \\
\hline
 & & & & 2X & -5 &
\end{array}$$

In this example therefore $q = X^2 - X - 1$, $r = 2X - 5$. ∎

The theorem on division with remainder allows one to find simple representants in $R[X]$ or residue classes modulo a principal ideal $(g) \subset R[X]$, provided that the leading coefficient of g is a unit. This is analogous to the representants $0, 1, \ldots, n-1$ modulo the ideal $n\mathbb{Z} \subset \mathbb{Z}$.

III.3.4 Theorem. *Suppose R is a ring with 1 and $g \in R[X]$ satisfies $\deg(g) > 0$ and the leading coefficient of g is a unit in R.*

The following is a bijection between sets:

$$\left\{h \in R[X] :\ \deg(h) < \deg(g)\right\} \longrightarrow R[X]/(g) \qquad h \mapsto h + (g).$$

In other words, every residue class modulo the ideal (g) has a unique representant $h \in R[X]$ with $\deg(h) < n$.

Proof. We first show that the given map is surjective. Let $f + (g)$ with $f \in R[X]$ be any residue class modulo the ideal (g). We apply Theorem III.3.1 to these f and g, obtaining $q, r \in R[X]$ with $\deg(r) < \deg(g)$. Observe that

$$f = qg + r \implies f + (g) = r + qg + (g) = r + (g),$$

since $qg \in (g)$. So indeed the map is surjective. To verify injectivity we use

$$h_1 + (g) = h_2 + (g) \iff h_1 - h_2 \in (g) \iff h_1 - h_2 = fg \text{ for some } f \in R[X].$$

The leading coefficient of g being a unit, one has $\deg(fg) = \deg(f) + \deg(g)$. So if $h_1 + (g) = h_2 + (g)$ and $\deg(h_1), \deg(h_2) < \deg(g)$ then $h_1 - h_2 = fg$ is only possible when $f = 0$, so in case $h_1 = h_2$.

From surjectivity and injectivity it follows that the map is bijective, finishing the proof. ∎

III.3.5 Remark. The set on the left and the set on the right in Theorem III.3.4 are additive groups. For the set on the left this follows using

$$\deg(h_1 + h_2) \leq \max(\deg(h_1), \deg(h_2)) < \deg(g)$$

in case $\deg(h_1), \deg(h_2) < \deg(g)$. The set on the right is even a ring. The bijection given between these groups is in fact an isomorphism of additive groups, since $h_1 + (g) + h_2 + (g) = h_1 + h_2 + (g)$. In particular this shows that the additive group of $R[X]/(g)$ only depends on the degree of g. In fact this additive group $R[X]/(g)$ is isomorphic to R^m, with $m = \deg(g)$, via

$$a_0 + a_1 X + \ldots + a_{m-1} X^{m-1} \bmod (g) \longmapsto (a_0, a_1, \ldots, a_{m-1}).$$

However, the product in the ring $R[X]/(g)$ does depend on the polynomial g, as is illustrated in the example below.

III.3.6 Example. Take $g = X^2 + X + 1 \in \mathbb{F}_2[X]$ and write $0 := \bar{0}$, $1 := \bar{1}$ for the elements of $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Since any polynomial in $\mathbb{F}_2[X]$ has coefficients $\{0, 1\}$, Theorem III.3.4 yields exactly 4 representants of the residue classes modulo (g):

$$0, \quad 1, \quad X, \quad X + 1.$$

Writing

$$x := X + (g) \qquad \text{one obtains} \quad \mathbb{F}_2[X]/(g) = \{0, 1, x, x + 1\}.$$

The additive group of $\mathbb{F}_2[X]/(g)$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ via $ax + b \mapsto (a, b)$. The product is slightly more complicated. In $\mathbb{F}_2[X]/(g)$ we have $x^2 + x + 1 = 0$ since $X^2 + X + 1 = g \in 0 + (g)$. As a consequence,

$$x(x + 1) = x^2 + x = 1 \cdot (x^2 + x + 1) + 1 = 1.$$

(for the second equality one uses division with remainder (!)). Therefore, in the ring $\mathbb{F}_2[X]/(g)$ $x + 1$ and x are each other's inverse. Moreover $1 \cdot 1 = 1$, hence every nonzero element has an inverse. We conclude that $\mathbb{F}_2[X]/(X^2 + X + 1)$ is a field consisting of 4 elements!

In contrast, the ring $\mathbb{F}_2[X]/(X^2 + X)$ is *not* a field. The representants of the residue classes are the same as before, however here $X(X + 1) \in 0 + (X^2 + X)$, hence the residue classes $X + (X^2 + X)$ and $X + 1 + (X^2 + X)$ are zero divisors in $\mathbb{F}_2[X]/(X^2 + X)$.

In the ring $\mathbb{F}_2[X]/(X^2)$ the element $X + (X^2)$ is even nilpotent, as follows from $(X + (X^2))^2 = X^2 + (X^2) = 0 + (X^2)$. This ring is therefore not isomorphic to $\mathbb{F}_2[X]/(g)$ nor to $\mathbb{F}_2[X]/(X^2 + X)$ (the latter ring has no nilpotent elements, in fact $r^2 = r$ for all $r \in \mathbb{F}_2[X]/(X^2 + X)$ as one easily verifies). ∎

III.3.7 Example. Let

$$S^1 := \{(a, b) \in \mathbb{R}^2 : \ a^2 + b^2 = 1\},$$

denote the circle in $\mathbb{R}^2$ with radius 1 and center $(0, 0)$. The ring consisting of all continuous functions from S^1 to $\mathbb{R}$ we denote by $C(S^1, \mathbb{R})$.

A polynomial $f \in \mathbb{R}[X,Y]$ yields a function $\rho_f : \mathbb{R}^2 \to \mathbb{R}$ (compare Remark III.2.5) given by $\rho_f(x,y) = f(x,y)$ for $(x,y) \in \mathbb{R}^2$. Restricting to $S^1 \subset \mathbb{R}^2$ one obtains a ring homomorphism

$$\Phi : \mathbb{R}[X,Y] \longrightarrow C(S^1,\mathbb{R}), \qquad f \mapsto \rho_f|_{S^1}.$$

The homomorphism Φ is not injective, for example $X^2+Y^2-1 \in \mathrm{Ker}(\Phi)$ since $a^2+b^2-1=0$ for all $(a,b) \in S^1$.

We will show:

$$\mathrm{Ker}(\Phi) = (X^2+Y^2-1) = (X^2+Y^2-1)\mathbb{R}[X,Y].$$

Proof. '$\supset$:' Every element in (X^2+Y^2-1) can be written as $(X^2+Y^2-1)f$, and $\Phi((X^2+Y^2-1)f) = \Phi(X^2+Y^2-1)\Phi(f) = 0 \cdot \Phi(f) = 0$.
'$\subset$:' Take $f \in \mathrm{Ker}(\Phi)$, so $f(a,b) = 0$ for all $(a,b) \in S^1$. We divide f by X^2+Y^2-1 in the ring $(\mathbb{R}[X])[Y] = \mathbb{R}[X,Y]$. Since $\deg_Y(X^2+Y^2-1) = 2$ this yields $q, r \in \mathbb{R}[X,Y]$ such that

$$f = q(X^2+Y^2-1)+r, \qquad r = r_0 + r_1 Y,$$

and $r_0, r_1 \in \mathbb{R}[X]$. The assumption $f \in \mathrm{Ker}(\Phi)$ implies

$$r(a,b) = r_0(a) + r_1(a)b = 0, \qquad \forall\,(a,b) \in S^1.$$

Note that if $(a,b) \in S^1$ then $(a,-b) \in S^1$ as well, hence we also have

$$r(a,-b) = r_0(a) - r_1(a)b = 0 \qquad \forall\,(a,b) \in S^1.$$

Adding and subtracting these equalities for fixed $a \in \mathbb{R}$ with $-1 < a < 1$ one finds

$$r_0(a) = r_1(a) = 0 \qquad \forall a \in \mathbb{R} \text{ such that } -1 < a < 1.$$

The polynomials $r_0, r_1 \in \mathbb{R}[X]$ therefore have infinitely many zeros, which implies $r_0 = r_1 = 0$ (see also Theorem III.5.2). It follows that $f = q(X^2+Y^2-1) \in (X^2+Y^2-1)$. This proves $\mathrm{Ker}(\Phi) = (X^2+Y^2-1)$.

The first isomorphism theorem II.3.7 now shows

$$\mathbb{R}[X,Y]/(X^2+Y^2-1) \cong \Phi(\mathbb{R}[X,Y]) \quad (\subset C(S^1,\mathbb{R})),$$

the image $\Phi(\mathbb{R}[X,Y])$ is sometimes called the ring of polynomial functions on the circle.

Theorem III.3.4 (applied to $R[Y]$ with $R = \mathbb{R}[X]$ and $g = Y^2+X^2-1 \in R[Y]$) allows an explicit description of this ring of polynomial functions on the circle. The representants in $\mathbb{R}[X,Y]$ of the residue classes modulo $I := (X^2+Y^2-1)$ are

$$f + gY \qquad f,g \in \mathbb{R}[X].$$

In $\mathbb{R}[X,Y]/I$ define the elements

$$x := \overline{X} = X + I, \quad y := \overline{Y} = Y + I.$$

Since $X^2+Y^2-1 \in I$, it follows that $x^2+y^2-1 = 0$, i.e., $y^2 = 1-x^2$. Summarizing:

$$\mathbb{R}[X,Y]/I = \{f + gy : f,g \in \mathbb{R}[x]\}$$

and such functions on the circle are multiplied using the rule

$$(f+gy)(h+ky) = (fh + gk\cdot(1-x^2)) + (fk+gh)y.$$

III.4 Rings of polynomials over a field

In this section we show that a polynomial ring $K[X]$ in one variable X over a field K is a principal ideal ring. As a consequence, the kernel of an evaluation homomorphism $\mathrm{ev}_s : K[X] \to S$ has the form (g) for some $g \in K[X]$.

III.4.1 Theorem. *Let K be a field. Every ideal in the ring $K[X]$ is a principal ideal.*

In case the ideal $I \subset K[X]$ is not the zero ideal, every polynomial $g \in I$, $g \neq 0$ of minimal degree is a generator of I, i.e., $I = (g)$.

Proof. Let $I \subset K[X]$ be an ideal. We must find $g \in I$ such that $I = K[X] \cdot g$. If $I = \{0\}$ then $g = 0$ works. Suppose $I \neq \{0\}$ and choose any $g \in I, g \neq 0$ such that $\deg(g)$ is as small as possible. We claim

$$I = K[X] \cdot g.$$

The inclusion $\supseteq$ is clear since $g \in I$ and I is an ideal, which implies $fg \in I$ for all $f \in K[X]$. To show the inclusion $\subseteq$, let $f \in I$. Since K is a field, the leading coefficient of g is a unit in K. Hence Theorem III.3.1 yields $q, r \in K[X]$ with

$$f = qg + r, \qquad \deg(r) < \deg(g).$$

Now $f, qg \in I$ and I is an additive group, so also $r = f - qg \in I$. Were $r \neq 0$ then r is a nonzero element in I of degree less than the degree of g, contradicting the minimal choice of g. This shows $r = 0$ and $f = qg \in K[X]g$. Hence every $f \in I$ is in $K[X] \cdot g$, showing $\subseteq$. This finishes the proof of Theorem III.4.1. ∎

III.4.2 Remark. The condition in Theorem III.4.1 that K should be a field is essential. For example the ideal $(2, X) \subset \mathbb{Z}[X]$ is not principal, see Exercise 16 on page 32.

Also a polynomial ring such as $\mathbb{R}[X, Y]$ contains ideals that are not principal, for example (X, Y) as shown in Example II.4.3.

III.4.3 Example. Using $\mathbb{R} \subset \mathbb{C}$ one obtains the evaluation homomorphism

$$\mathrm{ev}_i : \mathbb{R}[X] \longrightarrow \mathbb{C}, \qquad f \mapsto f(i).$$

Note that ev_i is surjective. As $i \notin \mathbb{R}$ no polynomials $\neq 0$ of degree ≤ 1 are in $\mathrm{Ker}(\mathrm{ev}_i)$. And $i^2 = -1$ shows that $g := X^2 + 1 \in \mathrm{Ker}(\Phi_i)$, with $\deg(g) = 2$. So g is a nonzero polynomial of minimal degree in $\mathrm{Ker}(\Phi_i)$ and therefore Theorem III.4.1 implies $\mathrm{Ker}(\mathrm{ev}_i) = (g)$. Furthermore it follows from the first isomorphism theorem II.3.7 that

$$\mathbb{R}[X]/(X^2 + 1) \cong \mathbb{C}.$$

For generalisations of this we refer to Exercise 12 on page 49. ∎

III.4.4 Example. Let V be a vector space over a field K and let $A \in \mathrm{End}_K(V)$. Then the kernel of the evaluation homomorphism

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

(see Example III.2.4) is a principal ideal (g) in $K[X]$. Write $a_n \in K - \{0\}$ for the leading coefficient of g, then $(a_n^{-1} g) = (g)$ and $a_n^{-1} g$ is a monic polynomial. Moreover by the minimality of the degree, $\mathrm{Ker}(\mathrm{ev}_A)$ contains only one monic polynomial of this minimal degree (otherwise the difference of two such monic polynomials would be an element of the kernel of still smaller degree). The *minimal polynomial* of A is by definition the *monic* polynomial $m_A \in K[X]$ such that

$$\mathrm{Ker}(\mathrm{ev}_A) = (m_A) = K[X] \cdot m_A.$$

In case $A = \lambda I$ with $\lambda \in K - \{0\}$ one finds $m_A = X - \lambda$. Take $\lambda, \mu \in K$ and put

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \qquad f := (X-\lambda)(X-\mu).$$

If $\lambda = \mu$ then $A = \lambda I$ hence $m_A = X - \lambda$. If $\lambda \neq \mu$ then $A - \tau I \neq 0$ for every $\tau \in K$, so the minimal polynomial has degree ≥ 2. We compute $f(A) = (A-\lambda)(A-\mu) =$

$$\left(\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) \cdot \left(\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} - \begin{pmatrix} \mu & 0 \\ 0 & \mu \end{pmatrix}\right)$$
$$=$$
$$\begin{pmatrix} 0 & 0 \\ 0 & \mu - \lambda \end{pmatrix} \begin{pmatrix} \lambda - \mu & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

Since $f(A) = 0$ and $\deg(f) = 2$ the conclusion is that f is the monic polynomial of minimal degree in $\mathrm{Ker}(\mathrm{ev}_A)$, so $\mathrm{Ker}(\mathrm{ev}_A) = (f)$ and $m_A = f$.

More generally let

$$A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n), \qquad \text{with} \quad \lambda_i \neq \lambda_j \text{ for } i \neq j,$$

i.e., A is a diagonal matrix with coefficients $A_{ii} = \lambda_i,\ A_{ij} = 0$ if $i \neq j$. Take

$$f = (X-\lambda_1)(X-\lambda_2)\ldots(X-\lambda_n).$$

Since $\mathrm{ev}_A(f) = f(A) = 0$ (verify!) we have $f \in (m_A)$. Hence $f = g m_A$ for some $g \in K[X]$. If $f \neq m_A$ then at least one of the factors $(X - \lambda_i)$ of f does not appear in m_A. This would contradict $m_A(A) = 0$ as one easily calculates. The conclusion is that $f = m_A$. Try to determine m_A for yourself in the case that some of the $\lambda_i$coincide. ∎

III.5 Rings of polynomials over a domain

III.5.1 Theorem. *Let R be a domain and $f \in R[X]$. Suppose $\alpha_1, \alpha_2, \ldots, \alpha_n \in R$ are pairwise distinct zeros of f. Then $q \in R[X]$ exists such that*

$$f = q \cdot (X-\alpha_1) \cdot (X-\alpha_2)\ldots(X-\alpha_n).$$

Proof. We use induction w.r.t. n. For $n = 1$ apply Theorem III.3.1:

$$f = q \cdot (X-\alpha_1) + r \qquad \text{with} \quad \deg(r) \leq 0,$$

so r is a constant. Now apply the evaluation homomorphism ev_{α_1}; this is possible since the domain R is by definition commutative. One obtains $0 = f(\alpha_1) = q(\alpha_1)\cdot 0 + r$ hence $r = 0$, proving the result in case $n = 1$.

Now let $n > 1$ and assume (induction hypothesis) the result for polynomials of degree $< n$. From $f(\alpha_n) = 0$ and division with remainder (as in the case $n = 1$) one finds

$$f = f_1 \cdot (X - \alpha_n).$$

For $1 \leq i \leq n-1$ we have

$$f_1(\alpha_i) \cdot (\alpha_i - \alpha_n) = f(\alpha_i) = 0,$$

hence since $\alpha_i \neq \alpha_n$ for $i < n$ and R is a domain it follows that

$$f_1(\alpha_i) = 0 \qquad (1 \leq i \leq n-1).$$

The induction hypothesis applied to f_1 shows

$$f_1 = q \cdot (X - \alpha_1)(X - \alpha_2) \ldots (X - \alpha_{n-1})$$

for some $q \in R[X]$, and therefore

$$f = f_1 \cdot (X - \alpha_n) = q \cdot (X - \alpha_1)(X - \alpha_2) \ldots (X - \alpha_n),$$

finishing the proof of Theorem III.5.1. ∎

III.5.2 Theorem. *If R is a domain and $f \in R[X]$ a nonzero polynomial, then the number of pairwise distinct zeros of f in R is at most* $\deg(f)$.

Proof. This follows from Theorem III.5.1: if $\alpha_1, \ldots, \alpha_n \in R$ are pairwise distinct zeros of f then $f = q \cdot (X - \alpha_1) \ldots (X - \alpha_n)$. Comparing degrees shows $\deg(f) = \deg(q) + n$, hence $\deg(f) \geq n$ since $q \neq 0$. ∎

III.5.3 Remark. The condition that R is a domain is essential. The polynomial $X^2 - \bar{1}$ in $(\mathbb{Z}/8\mathbb{Z})[X]$ has degree 2 while it has 4 zeros in $\mathbb{Z}/8\mathbb{Z}$, namely $\bar{1}, \bar{3}, \bar{5}, \bar{7}$.

In the non-commutative division algebra of the quaternions $\mathbb{H}$, see I.1.5, the polynomial $X^2 + 1 \in \mathbb{H}[X]$ has degree 2 and for example the zeros $\pm i, \pm j, \pm k$ (in fact this polynomial has infinitely many zeros in $\mathbb{H}$, see Exercise 4 on page 49).

Using a result on finite abelian groups, a consequence of Theorem III.5.2 is the following.

III.5.4 Corollary. *If R is a domain and $G \subset R^\times$ a finite subgroup of $R^\times$, then G is cyclic, i.e., $g \in G$ exists with* $\operatorname{ord}(g) = \#G$.

Proof. By assumption G is a finite abelian group. If $\#G = 1$ the result is trivial. If $\#G > 1$, a result from Group Theory is that

$$G \cong (\mathbb{Z}/d_1\mathbb{Z}) \times (\mathbb{Z}/d_2\mathbb{Z}) \times \ldots \times (\mathbb{Z}/d_t\mathbb{Z})$$

for integers $2 \leq d_1 \leq d_2 \leq \ldots \leq d_t$ with (in case $t > 1$) $d_1 | d_2$, $d_2 | d_3$, …, $d_{t-1} | d_t$. In particular $\#G = \prod d_j$ and every element of G has order dividing d_t. This means that all elements of G are zeros of the polynomial $X^{d_t} - 1 \in R[X]$. Therefore Theorem III.5.2 implies $\#G \leq d_t$. As a consequence $t = 1$ and G is cyclic, proving the result. ∎

III.6 Differentiation

In Exercise 26 on page 33 the ring of dual numbers over a field was introduced. We now generalise this to other rings, and use it to introduce and study the 'derivative' of a polynomial over a ring.

III.6.1 Definition. Given a ring R, the *ring of dual numbers* over R consists of all formal expressions

$$r + s \cdot \epsilon, \qquad r, s \in R$$

with addition $(r + s \cdot \epsilon) + (r' + s' \cdot \epsilon) = (r + r') + (s + s') \cdot \epsilon$ and multiplication law given by $(r + s \cdot \epsilon) \cdot (r' + s' \cdot \epsilon) = rr' + (rs' + sr') \cdot \epsilon$. This ring is denoted by $R[\epsilon]$.

One verifies without difficulty (regardless of R being commutative or not) that indeed $R[\epsilon]$ is a ring. We consider R as a subring of $R[\epsilon]$ by identifying $r \in R$ with $r + 0 \cdot \epsilon \in R[\epsilon]$. By definition, $r + s \cdot \epsilon = r' + s' \cdot \epsilon$ if and only if $r = r'$ and $s = s'$, for all $r, r', s, s' \in R$. An element of $R[\epsilon]$ of the form $0 + s \cdot \epsilon$ will simply be written as $s \cdot \epsilon$. Note that $(s \cdot \epsilon)^2 = 0$ whenever $s \in R$.

Let R be a ring with 1. The dual numbers over the polynomial ring $R[X]$ are denoted, as a special case of the above, by $R[X][\epsilon]$. Note that $X + \epsilon := X + 1 \cdot \epsilon$ commutes with all $f + g \cdot \epsilon \in R[X][\epsilon]$, by definition of the multiplication of polynomials and by definition of the multiplication of dual numbers.

III.6.2 Lemma. *If R is a ring with 1 which we consider as a subring of $R[X][\epsilon]$, then the evaluation map*

$$\mathrm{ev}_{X+\epsilon} : R[X] \longrightarrow R[X][\epsilon]$$

is a ring homomorphism. Writing the image of $f \in R[X]$ as $f(X + \epsilon) = f_0 + f_1 \cdot \epsilon$ with $f_0, f_1 \in R[X]$, one has $f_0 = f$.

Proof. Observing that $X + \epsilon$ commutes with the elements of R, Theorem III.2.1 implies that $\mathrm{ev}_{X+\epsilon}$ is a ring homomorphism.

As is easily verified, for any ring S the map $S[\epsilon] \to S$ given by $s_0 + s_1 \cdot \epsilon \mapsto s_0$ (this is the same as: substitute 0 for ϵ) is a ring homomorphism. In the case $S = R[X]$, clearly it sends $f(X + \epsilon) = f_0 + f_1 \cdot \epsilon$ to f and also to f_0, showing that $f = f_0$. ∎

III.6.3 Definition. Let R be a ring with 1 and $f \in R[X]$. Writing

$$f(X + \epsilon) = f + f_1 \cdot \epsilon \in R[X][\epsilon],$$

the *derivative* of f is the polynomial f_1. This is denoted f' or $\frac{df}{dX}$ or $\frac{d}{dX} f$ or, in case f may also be considered as a polynomial in a variable different from X, by $\frac{\partial f}{\partial X}$.

III.6.4 Remark. Note that this definition of 'derivative' does not involve limits. We have the formula

$$f(X + \epsilon) = f + f' \cdot \epsilon,$$

which holds in $R[X][\epsilon]$ for an arbitrary ring R with 1. Note the analogy with a 'first order Taylor expansion' as discussed in Calculus or Analysis courses.

III.6.5 Theorem. *Let R be a ring with 1.*

(a) For all $f, g \in R[X]$ we have

$$(f + g)' = f' + g' \quad \textit{and} \quad (fg)' = f'g + fg'.$$

(b) If $f = \sum_{k=0}^{n} a_k X^k \in R[X]$ then

$$f' = \sum_{k=1}^{n} k a_k X^{k-1}.$$

(Here $k a_k = a_k + a_k + \ldots + a_k$ (k terms).)

Proof. (a) follows using that $\mathrm{ev}_{X+\epsilon}$ is a ring homomorphism: by definition we have $f(X + \epsilon) = f + f' \cdot \epsilon$ and $g(X + \epsilon) = g + g' \cdot \epsilon$. Hence

$$(f + g) + (f + g)' \cdot \epsilon = \mathrm{ev}_{X+\epsilon}(f + g) = (f + f' \cdot \epsilon) + (g + g' \cdot \epsilon) = (f + g) + (f' + g') \cdot \epsilon,$$

showing that $(f + g)' = f' + g'$. Similarly

$$(fg) + (fg)' \cdot \epsilon = \mathrm{ev}_{X+\epsilon}(fg) = (f + f' \cdot \epsilon)(g + g' \cdot \epsilon) = (fg) + (fg' + g'f) \cdot \epsilon,$$

which shows $(fg)' = fg' + f'g$.

To prove (b), one first observes that

$$(a \cdot X^k)' = kaX^{k-1} \quad \text{for } a \in R, k \in \mathbb{Z}_{>0}.$$

This is immediate from $a \cdot (X+\epsilon)^k = aX^k + kaX^{k-1} \cdot \epsilon$, using either Newton's Binomium (compare Exercise 15 on page 14; this is valid here because X and ϵ commute), or by using a straightforward induction with respect to k. We also have $(a)' = 0$ for $a \in R$ (verify!). Hence using (a) one obtains

$$(\sum_{k=0}^{n} a_k X^k)' = \sum_{k=0}^{n} (a_k X^k)' = \sum_{k=1}^{n} k a_k X^{k-1},$$

finishing the proof of Theorem III.6.5. ∎

III.6.6 Remark. Alternatively, we could have used the formula in III.6.5(b) to define the derivative f'. In that case the properties stated in Theorem III.6.5(a) would require a different proof.

For us, the most important application of the derivative will be the study of *multiple zeros* of polynomials. Suppose R is a commutative ring. If $\alpha \in R$ is a zero of $f \in R[X]$, then using the proof of Theorem III.5.1 $f = (X-\alpha) \cdot q$, with $q \in R[X]$. If we can even write $f = (X-\alpha)^2 \cdot q_1$ with $q_1 \in R[X]$ (so in case R is a domain this means α is a zero of q as well), then α is called a *double* or *multiple* zero of f.

III.6.7 Theorem. *Let R be a commutative ring with 1 and let $f \in R[X]$. Suppose $\alpha \in R$ is a zero of f. Then: α is a double zero of $f \iff \alpha$ is a zero of f'.*

Proof. Write $f = (X-\alpha) \cdot q$, with $q \in R[X]$. Then

$$\alpha \text{ is a double zero of } f \iff q(\alpha) = 0.$$

Using $f = (X-\alpha) \cdot q$ and III.6.5(a) it follows that

$$f' = (X-\alpha)' \cdot q + (X-\alpha) \cdot q' = q + (X-\alpha)q'.$$

As a consequence

$$f'(\alpha) = q(\alpha).$$

This shows $q(\alpha) = 0 \Leftrightarrow f'(\alpha) = 0$, proving Theorem III.6.7. ∎

III.7 Exercises

1. Let R be a ring without zero divisors, and take $f, g \in R[X]$.
   Show that $\deg(f \cdot g) = \deg(f) + \deg(g)$. Conclude that $R[X]$ has no zero divisors.
2. Show using the example $f = X^2$ and $g = 2X$ in the ring $\mathbb{Z}[X]$ that in Theorem III.3.1 the condition that the leading coefficient of g should be a unit, can not be missed.
3. Let K be a field, $f \in K[X]$, and $\alpha_0, \alpha_1, \ldots \alpha_n$ an $n+1$-tuple of pairwise different elements of K, with $n \geq \deg(f)$. Prove that
$$f = \sum_{i=0}^{n} f(\alpha_i) \frac{\prod_{j=0, j\neq i}^{n} (X - \alpha_j)}{\prod_{j=0, j\neq i}^{n} (\alpha_i - \alpha_j)},$$
   the interpolation formula of Lagrange.
4. Let $x = a + bi + cj + dk \in \mathbb{H}$, with $a, b, c, d \in \mathbb{R}$. Prove:
$$x \text{ is a zero of } X^2 + 1 \Leftrightarrow (x\overline{x} = 1 \text{ and } \overline{x} = -x) \Leftrightarrow (a = 0 \text{ and } b^2 + c^2 + d^2 = 1).$$
   Conclude that $X^2 + 1$ has infinitely many zeros in $\mathbb{H}$.
5. Suppose R is a commutative ring with 1 and let $f, g \in R[X]$ and $k \in \mathbb{Z}_{>0}$.
   (a) Show that $f \in R[X] \cdot g^k \implies f' \in R[X] \cdot g^{k-1}$.
   (b) Give an example showing that conversely $f' \in R[X] \cdot g^{k-1}$ does not necessarily imply that $f \in R[X] \cdot g^k$.
6. Let $R = \mathbb{F}_2$ and $f \in R[X]$.
   (a) Show that the next three conditions are equivalent:
      (i) $f' = 0$;
      (ii) f can be written as $f = \sum_{k=0}^{n} a_k X^{2k}$ with all $a_k \in \mathbb{F}_2$;
      (iii) $g \in \mathbb{F}_2[X]$ exists with $f = g^2$.
   (b) Show that $(f')' = 0$.
7. Let R be a ring with 1. For $f \in R[X]$ and $k \in \mathbb{Z}_{\geq 0}$ one defines $f^{(k)}$ inductively by $f^{(0)} = f$, $f^{(k)} = (f^{(k-1)})'$. Show that for all $f, g \in R[X]$ and $n \in \mathbb{Z}_{\geq 0}$ one has:
$$(f \cdot g)^{(n)} = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) f^{(k)} g^{(n-k)}$$
   (this is called the *formula of Leibniz*).
8. Let R be a finite ring. Show that $n, m \in \mathbb{Z}$ exist with $n > m > 0$ such that $x^n = x^m$ for all $x \in R$.
9. Let R be a domain and $f, g \in R[X]$ such that $\max\{\deg(f), \deg(g)\} < \#R$ (for example, this holds whenever R is infinite). Prove: $(\forall x \in R : f(x) = g(x)) \Leftrightarrow f = g$.
10. Let p be prime and $f, g \in \mathbb{F}_p[X]$. Prove:
$$(\forall x \in \mathbb{F}_p : f(x) = g(x)) \Leftrightarrow f - g \in \mathbb{F}_p[X] \cdot (X^p - X).$$
11. Define the evaluation homomorphism
$$\mathrm{ev} : \mathbb{R}[X, Y] \longrightarrow \mathbb{R}[T], \qquad f(X, Y) \mapsto f(T^2, T^3).$$
    Show that $\mathrm{Ker}(\mathrm{ev}) = (X^3 - Y^2)$ and that $\mathrm{ev}(\mathbb{R}[X, Y]) = \{\sum a_i T^i : a_1 = 0\}$.
12. (a) Take $z = a + bi \in \mathbb{C}$ with $z \notin \mathbb{R}$. Prove that the evaluation homomorphism
$$\mathrm{ev}_z : \mathbb{R}[X] \longrightarrow \mathbb{C}, \qquad f \mapsto f(z),$$
    (here we use the inclusion $\mathbb{R} \subset \mathbb{C}$) is surjective.

(b) Put $g = X^2 - 2aX + a^2 + b^2$. Show that

$$\mathrm{Ker}(\mathrm{ev}_z) = (g) \qquad \text{and} \qquad \mathbb{R}[X]/(g) \cong \mathbb{C}.$$

(c) Let $f = aX^2 + bX + c \in \mathbb{R}[X]$ with $a \neq 0$. Show:

$$\begin{array}{rcll} \mathbb{R}[X]/(f) & \cong & \mathbb{C} & \text{if} \quad b^2 - 4ac < 0, \\ & \cong & \mathbb{R}[\epsilon] & \text{if} \quad b^2 - 4ac = 0, \\ & \cong & \mathbb{R} \times \mathbb{R} & \text{if} \quad b^2 - 4ac > 0. \end{array}$$

Here $\mathbb{R}[\epsilon]$ is the ring of dual numbers over $\mathbb{R}$, see Definition III.6.1. Try to construct explicit isomorphisms in all three cases.

13. Take $z, w \in \mathbb{C} - \mathbb{R}$ and let

$$\mathrm{ev}_{z,w} : \mathbb{R}[X,Y] \longrightarrow \mathbb{C}, \qquad f \mapsto f(z,w),$$

be the evaluation homomorphism. Show that $\mathrm{Ker}(\mathrm{ev}_{z,w})$ is generated by one linear polynomial and one polynomial of degree 2. Determine such polynomials explicitly in case $z = 1 + i,\ w = 3 - 2i$.

14. (a) Verify that the tangent line to the circle

$$S^1 := \{(a,b) \in \mathbb{R}^2 : a^2 + b^2 = 1\}$$

in the point $(a,b) \in S^1$ is given by

$$\ell_{(a,b)}(X,Y) = 0,$$

where

$$\ell_{(a,b)} = a(X-a) + b(Y-b) \in \mathbb{R}[X,Y].$$

(b) Let $\mathbb{R}[\epsilon]$ be the ring of dual numbers over $\mathbb{R}$ (see Definition III.6.1). Define

$$S^1(\mathbb{R}[\epsilon]) := \left\{(a + s\epsilon,\, b + t\epsilon) \in \mathbb{R}[\epsilon]^2 :\ (a+s\epsilon)^2 + (b+t\epsilon)^2 = 1\right\},$$

'the points of S^1 with coordinates in $\mathbb{R}[\epsilon]$'. Show that:

$$(a + s\epsilon,\, b + t\epsilon) \in S^1(\mathbb{R}[\epsilon]) \iff (a,b) \in S^1 \text{ and } \ell_{(a,b)}(a+s, b+t) = 0.$$

15. Let K be a field and consider $R = K[X]/(X^n)$ for $n \in \mathbb{Z}_{\geq 1}$. Writing $x := X + (X^n) \in R$, every element r in R is of the form

$$r = a_0 + a_1 x + \ldots a_{n-1} x^{n-1}, \qquad a_i \in K.$$

(a) Show that $r \in R$ is a unit if and only if $a_0 \neq 0$. Find the inverse of such a unit.

(b) Show that every zero divisor in R is nilpotent. What is the smallest k such that $r^k = 0$ for all zero divisors r in R?

(c) Given any $a \in K$, find a ring isomorphism

$$K[X]/\left((X-a)^n\right) \cong K[X]/(X^n).$$

(d) Given any $n > 1$, find $f \in K[X]$ such that $f + (X^n)$ is a unit in R whereas $f + (X-1)^n \in K[X]/((X-1)^n)$ is nilpotent.

16. Let V be a finite dimensional vector space over a field K and let $A \in \mathrm{End}_K(V)$. Suppose $\lambda \in K$ is an eigenvalue of A, i.e., $v \in V$ exists with $v \neq 0$ and $Av = \lambda v$ (or equivalently, $A - \lambda$ is not invertible in the ring $\mathrm{End}_K(V)$).

(a) Show that λ is a zero of m_A. (Hint: consider $m_A(A)v$).

(b) Prove that every zero μ of m_A is an eigenvalue of A. (Hint: write $m_A = (X-\mu)\cdot g$ and consider $0 = (A-\mu)\cdot g(A)$.)

(c) Prove that in case A has $n = \dim_K(V)$ pairwise distinct eigenvalues in K, then m_A is, possibly up to a sign ± 1, equal to the characteristic polynomial $\det(A - XI)$ of A.

17. Determine the minimal polynomials of the following matrices in $\mathrm{M}_3(K)$, where K is a field. Distinguish the cases $\lambda = \mu$ and $\lambda \neq \mu$.

$$A := \begin{pmatrix} \lambda & 0 & 1 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{pmatrix}, \qquad B := \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{pmatrix}.$$

18. Let R be the ring of polynomial functions on the circle (compare Example III.3.7):

$$R = \mathbb{R}[X,Y]/I, \qquad \text{where} \quad I = (X^2 + Y^2 - 1).$$

Define $x, y \in R$ by $x := X + I$, $y := Y + I$ and consider the ideal $M := (x-1, y) \subset R$. Recall that every $r \in R$ can be given uniquely as $r = f + gy$ with $f, g \in \mathbb{R}[x]$.

(a) Prove that

$$\mathrm{ev}_{(1,0)} : R \longrightarrow \mathbb{R}, \qquad f + gy \mapsto f(1)$$

is a surjective ring homomorphism, and $\mathrm{Ker}(\mathrm{ev}_{(1,0)}) = M$.

(b) Define

$$N : R \longrightarrow \mathbb{R}[x]$$

by

$$N(f + gy) := (f + gy)(f - gy) = f^2 - g^2(1 - x^2).$$

Show that if $N(r) \in \mathbb{R}[x]$ is a constant, then $r = f + gy$ with f constant and $g = 0$.

(c) Prove that M is not a principal ideal. (Hint: suppose $1 - x = \alpha \cdot r$, $y = \alpha \cdot s$, and consider $N(1-x)$, $N(y)$.)

19. We define the n-sphere by

$$S^n := \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} : \; x_0^2 + x_1^2 + \ldots x_n^2 = 1\}.$$

Let $C(S^n, \mathbb{R})$ denote the ring of continuous functions from S^n to $\mathbb{R}$. Let Φ_n be the restriction map

$$\Phi_n \; : \mathbb{R}[X_0, X_1, \ldots, X_n] \longrightarrow C(S^1, \mathbb{R}), \qquad f \mapsto f_{|S^n}.$$

Show that

$$\mathrm{Ker}(\Phi_n) = (X_0^2 + X_1^2 + \ldots + X_n^2 - 1).$$

IV PRIME IDEALS AND MAXIMAL IDEALS

IV.1 Prime ideals

Throughout this chapter R is a *commutative* (unitary) ring.
An important property of prime numbers p is that

$$p|ab \implies p|a \text{ of } p|b$$

whenever $a, b \in \mathbb{Z}$. In other words:

$$ab \in p\mathbb{Z} \implies a \in p\mathbb{Z} \text{ or } b \in p\mathbb{Z}.$$

Ideals with this property will be called *prime ideals*:

IV.1.1 Definition. Let R be a commutative ring with 1. A *prime ideal* of R is an ideal $I \subset R$ satisfying:

(P1) $I \neq R$;
(P2) For all $a, b \in R$ such that $ab \in I$ it holds that either $a \in I$ or $b \in I$ (or both).

IV.1.2 Example. As we saw, $p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$ for every prime number p. If $n \in \mathbb{Z}_{>0}$ is not prime then $n\mathbb{Z}$ is *not* a prime ideal of $\mathbb{Z}$: namely, in case $n = 1$ condition (**P1**) is not satisfied, and if $n > 1$ then with $n = ab$ and $1 < a, b < n$ one obtains $ab = n \in n\mathbb{Z}$ although $a \notin n\mathbb{Z}$ and $b \notin n\mathbb{Z}$. Hence for $n > 1$ not prime $n\mathbb{Z}$ does not satisfy (**P2**).

The ideal $\{0\} \subset \mathbb{Z}$ is a prime ideal. ▬■

IV.1.3 Theorem. *The ideal $\{0\} \subset R$ is a prime ideal if and only if R is a domain.*

Proof. If R is a domain then $1 \neq 0$ hence $\{0\} \neq R$. Moreover in a domain $ab = 0 \Rightarrow a = 0$ or $b = 0$ so condition (**P2**) is satisfied.

Vice versa, is $\{0\}$ a prime ideal then (**P2**) implies that R has no zero divisors. Moreover $1 \in R$ is not an element of the prime ideal $\{0\}$ since otherwise $\{0\} = R$, contradicting $\{0\}$ being a prime ideal. So $1 \neq 0$ in R, which finishes the proof that R is a domain. This shows IV.1.3. ■

IV.1.4 Example. The ideal $\mathbb{R}[X] \cdot (X^2 - 1) \subset \mathbb{R}[X]$ is not a prime ideal, because it contains $(X + 1)(X - 1)$, but not $X + 1$ or $X - 1$.

However, the ideal $\mathbb{R}[X] \cdot (X^2 + 1) \subset \mathbb{R}[X]$ *is* a prime ideal. To verify this, we will use the evaluation homomorphism (see III.4.3):

$$\mathrm{ev}_i : \mathbb{R}[X] \longrightarrow \mathbb{C}, \qquad f \mapsto f(i).$$

As we observed in Example III.4.3,

$$\mathrm{Ker}(\mathrm{ev}_i) = \mathbb{R}[X]\cdot(X^2+1).$$

As a consequence, given $f, g \in \mathbb{R}[X]$ we have

$$fg \in \mathbb{R}[X]\cdot(X^2+1) \Rightarrow (fg)(i) = f(i)g(i) = 0 \Rightarrow f(i) = 0 \text{ or } g(i) = 0$$

$$\Rightarrow f \in \mathbb{R}[X]\cdot(X^2+1) \text{ or } g \in \mathbb{R}[X]\cdot(X^2+1).$$

This verifies (**P2**); we leave (**P1**) as an exercise to reader. ∎

The next result generalises IV.1.3, and asserts that one can verify whether an ideal I is prime by considering the residue class ring R/I.

IV.1.5 Theorem. *Let R be a commutative ring with 1, and $I \subset R$ an ideal. Then*

$$I \text{ is a prime ideal of } R \iff R/I \text{ is a domain.}$$

Proof. For $a \in R$ put $\overline{a} = (a+I) \in R/I$. The ring R/I is by Definition I.2.13 a domain if and only if $\overline{1} \neq \overline{0}$ and R/I has no zero divisors. Now

$$\overline{1} \neq \overline{0} \iff 1 \notin I \iff I \neq R \iff (\mathbf{P1}) \text{ holds,}$$

and

$$\begin{aligned} & R/I \text{ has no zero divisors} \\ \iff\ & (\forall\, \overline{a}, \overline{b} \in R/I : \overline{a}\overline{b} = \overline{0} \Rightarrow \overline{a} = \overline{0} \text{ or } \overline{b} = \overline{0}) \\ \iff\ & (\forall\, a, b \in R : ab \in I \Rightarrow a \in I \text{ or } b \in I) \\ \iff\ & (\mathbf{P2}) \text{ holds.} \end{aligned}$$

Here we repeatedly used that $\overline{c} = \overline{0}$ is equivalent to $c \in I$. So we conclude R/I is a domain $\Leftrightarrow$ (**P1**) and (**P2**) hold $\Leftrightarrow$ I is a prime ideal of R. This proves IV.1.5. ∎

IV.1.6 Example. We know $\mathbb{R}[X]/(X^2+1) \cong \mathbb{C}$, which is a domain. Using Theorem IV.1.5 this implies that (X^2+1) is a prime ideal of $\mathbb{R}[X]$. ∎

In many cases a fast way to verify whether some ideal $I \subset R$ is prime consists of calculating the ring R/I and applying IV.1.5. For special rings, other methods exist: see for example Theorem V.2.4 below.

IV.1.7 Example. Take the ideals $J = (X+Y, X^2+X+Y+1)$ and $I = (X+Y) = (Y-(-X))$ in the ring $R = \mathbb{R}[X,Y]$. Then (see II.3.8)

$$R/I \cong \mathbb{R}[X], \qquad F(X,Y) \mapsto F(X,-X).$$

If $\phi : R \to R/I$ denotes the canonical map, then $\phi(J) = (0, X^2+X+(-X)+1) = (X^2+1) \subset \mathbb{R}[X]$ hence by II.3.10 one finds

$$R/J \cong \mathbb{R}[X]/(X^2+1) \cong \mathbb{C}.$$

Therefore J is a prime ideal of R. ∎

IV.1.8 Example. Consider the ideals $J = (5, X^2+Y+1)$ and $I = (Y-(-X^2-1)) \subset J$ in the ring $R = \mathbb{Z}[X,Y]$. Then

$$R/I \cong \mathbb{Z}[X], \qquad F(X,Y) \mapsto F(X,-X^2-1),$$

and $J/I = (5,0) = (5)$. As is not hard to verify,

$$\phi : \mathbb{Z}[X] \longrightarrow \mathbb{F}_5[X], \qquad \sum_i a_i X^i \mapsto \sum_i \overline{a_i} X^i,$$

(here $\overline{a_i} \in \mathbb{F}_5$) is a surjective ring homomorphism with $\mathrm{Ker}(\phi) = (5) = 5\mathbb{Z}[X]$ (see also II.3.9). Using II.3.7 and II.3.10 it follows that

$$R/J \cong \mathbb{F}_5[X].$$

The latter ring is a domain since $\mathbb{F}_5$ is a domain (even a field). As a consequence J is a prime ideal. ∎

IV.1.9 Example. Take $I = (YZ - X^2, X^2 - Z) \subset \mathbb{C}[X,Y,Z] = R$. Similar to the examples above, $R/I \cong \mathbb{C}[X,Y]/(YX^2 - X^2)$. From

$$X^2 \cdot (Y-1) \in (YX^2 - X^2)$$

and

$$X^2 \notin (YX^2 - X^2), \;\; Y - 1 \notin (YX^2 - X^2),$$

it follows that $(YX^2 - X^2)$ is *not* a prime ideal of $\mathbb{C}[X,Y]$. Hence R/I is not a domain, and therefore I is not a prime ideal of R. ∎

IV.2 Maximal ideals

IV.2.1 Definition. Let R be a commutative ring with 1. An ideal M of R is called *maximal* if

(M1) $M \neq R$;
(M2) for every ideal J of R with $M \subset J \subset R$ either $J = M$ or $J = R$.

So a maximal ideal 'cannot be made bigger' without obtaining the full ring.

Examples of ideals that are not maximal: $9\mathbb{Z} \subset \mathbb{Z}$, because ideal $3\mathbb{Z}$ is 'strictly between'; also $(2) \subset \mathbb{Z}[X]$, since $(2,X)$ is strictly between (2) and $\mathbb{Z}[X]$.

Examples of ideals that are maximal are readily found once we have shown the analog of IV.1.5 for maximal ideals. We start by showing the analog of IV.1.3:

IV.2.2 Theorem. *The ideal $\{0\} \subset R$ is maximal if and only if R is a field.*

Proof. $\Leftarrow$. In a field one has $1 \neq 0$, hence $\{0\} \neq R$, hence $\{0\}$ satisfies (**M1**). Moreover by II.4.5 a field contains no ideals except $\{0\}$ and R, hence (**M2**) is satisfied as well. This shows $\Leftarrow$.
$\Rightarrow$. We claim that every $a \in R, a \neq 0$ has an inverse. To obtain this we apply (**M2**) to the ideal $J = Ra$. This ideal is different from $\{0\}$, so by (**M2**) (with $M = \{0\}$) it follows that $Ra = R$. Hence $1 \in Ra$, so $1 = ba$ for some $b \in R$, which means a has an inverse. As R is unitary this means that R is a field. This shows $\Rightarrow$, and finishes the proof of IV.2.2. ∎

IV.2.3 Theorem. *If R is a commutative ring with 1 and $M \subset R$ an ideal, then:*

$$M \text{ is a maximal ideal of } R \iff R/M \text{ is a field.}$$

Proof. The idea of the proof is to reduce the assertion to the special case IV.2.2, by means of II.3.10.

Write $\overline{R} = R/M$. By II.3.10 the ideals J in R such that $M \subset J \subset R$ are in $1-1$ correspondence with the ideals $\overline{J} = J/M$ of $\overline{R}$. Therefore an R-ideal J strictly between M and R gives rise to an $\overline{R}$-ideal strictly between $\{\overline{0}\}$ and $\overline{R}$, and vice versa. As a result,

$$M \text{ is a maximal ideal of } R \iff \{\overline{0}\} \text{ is a maximal ideal of } \overline{R} = R/M.$$

By IV.2.2 this last condition is equivalent to: $\overline{R} = R/M$ is a field. This proves IV.2.3. ∎

Theorem IV.2.3 is used in a similar way to determine whether an ideal is maximal, as Theorem IV.1.5 is used to see whether an ideal is prime.

IV.2.4 Example. Take $n \in \mathbb{Z}_{>0}$. Then $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime (see I.2.11), hence

$$p\mathbb{Z} \subset \mathbb{Z} \text{ is maximal, for all } p \text{ prime}$$

and the ideals (n) with n not a prime number are not maximal. Indeed we have for $n = ab$ and $1 < a, b < n$ that $(n) \subset (a)$ and $(a) \neq \mathbb{Z}$ (since $a \neq \pm 1$). ∎

IV.2.5 Example. We saw in Example IV.1.8 that

$$\mathbb{Z}[X,Y]/(5, X^2 + Y + 1) \cong \mathbb{F}_5[X].$$

This is a domain but not a field ($X^{-1} \notin \mathbb{F}_5[X]$), so $(5, X^2 + Y + 1) \subset \mathbb{Z}[X,Y]$ is a prime ideal but it is not maximal. ∎

IV.2.6 Example. For every $(a, b) \in \mathbb{R} \times \mathbb{R}$ it holds that $(X - a, Y - b)$ is a maximal ideal in $\mathbb{R}[X, Y]$ because (see II.3.11)

$$\mathbb{R}[X,Y]/(X - a, Y - b) \cong \mathbb{R}[X]/(X - a) \cong \mathbb{R},$$

and $\mathbb{R}$ is a field. ∎

IV.2.7 Corollary. *Every maximal ideal is prime.*

Proof. This is immediate from IV.2.3 and IV.1.5, using that any field is a domain. ∎

IV.2.8 Remark. As Example IV.2.5 shows, the converse of IV.2.7 is false. An even simpler example showing this is $\{0\} \subset \mathbb{Z}$: the ring $\mathbb{Z}$ is a domain but not a field, so $\{0\} \subset \mathbb{Z}$ is prime but not maximal.

IV.3 Zorn's lemma

IV.3.1 Theorem. *Every commutative ring R with $1 \neq 0$ contains a maximal ideal.*

The idea of the proof of this is quite simple: start with the zero-ideal $\{0\}$, and keep enlarging this until no further enlargement is possible without obtaining the full ring. We will first treat a case where the proof along these lines can indeed be completed by ordinary means. It turns out that for the general case, a tool from set theory is necessary: Zorn's lemma (named after the German mathematician Max August Zorn, 1906–1993, although already before Zorn it was also treated by the Polish mathematician Kazimierz Kuratowski, 1896–1980).

Proof. (in the special case that R is countable.) Enumerate the elements of R as $r_1, r_2, \ldots$. Inductively define the sequence of ideals $I_0 \subseteq I_1 \subseteq \ldots$ by $I_0 = (0)$ and

$$I_n = \begin{cases} I_{n-1} + (r_n) & \text{if } I_{n-1} + (r_n) \neq R; \\ I_{n-1} & \text{otherwise.} \end{cases}$$

Put

$$M = \bigcup_{n \in \mathbb{N}} I_n.$$

We claim that M is a maximal ideal. To verify this, we first need to check that M is an ideal. Let $a,b \in M$, then $n,m \in \mathbb{N}$ exist with $a \in I_n$ and $b \in I_m$. Taking $\ell := \max\{n,m\}$ we find $I_n, I_m \subseteq I_\ell$, hence $a,b \in I_\ell$ which is an ideal. Therefore $a-b \in I_\ell \subseteq M$. Moreover given any $r \in R$ then also $ra \in I_n \subseteq M$ and this shows that M is an ideal of R.

If M is not a maximal ideal then one of the following two situations happens. Either the unit element of R is in M or some ideal N of R exists with $M \subsetneq N \subsetneq R$. If $1 \in M$ then $1 \in I_n$ for some n, hence this I_n equals R, contradicting the definition of the ideals I_n. In the remaining case $M \subsetneq N \subsetneq R$, there exists $r_n \in N$ with $r_n \notin M$. This implies $M + (r_n) \subseteq N$ hence $M + (r_n) \neq R$. Therefore certainly $I_{n-1} + (r_n) \neq R$, since $I_{n-1} \subset M$. The definition of I_n now shows that $r_n \in I_n \subset M$, a contradiction. This completes the proof that M is maximal, and thereby the proof of Theorem IV.3.1 in the special case that R is countable. ∎

It is a fact that in order to extend the argument presented above to more general rings, one needs an axiom from set theory called the 'Axiom of Choice', which is a surprisingly simple sounding statement:

Axiom of Choice. *If S,T are sets and $f : S \to T$ a surjective map, then $g : T \to S$ exists with $f \circ g = \mathrm{id}_T$.*

It is a standard but nontrivial result from set theory that Zorn's lemma, which will be stated below, is equivalent to the Axiom of Choice. In order to formulate Zorn's Lemma, two additional definitions are needed.

IV.3.2 Definition. A *partially ordered set* is a pair $(P,\leq)$ in which P is a set and $\leq$ is a binary relation on P with the properties

$$\begin{aligned} &\forall x,y,z \in P: \quad (x \leq y \wedge y \leq z) \Rightarrow x \leq z, \\ &\forall x,y \in P: \quad\quad (x \leq y \wedge y \leq x) \Leftrightarrow x = y. \end{aligned}$$

IV.3.3 Definition. A *chain* in a partially ordered set $(P,\leq)$ is a subset $K \subset P$ with the property

$$\forall x,y \in K: \quad x \leq y \ \vee \ y \leq x.$$

The chain K is called *maximal* if $K \subseteq L \subseteq P$ and L is a chain $\Longrightarrow L = K$.

In other words, a chain K in a partially ordered set $(P,\leq)$ is maximal if for every y in the complement of K in P, some $x \in K$ exists such that neither $x \leq y$ nor $y \leq x$. We are now ready to state Zorn's Lemma.

Zorn's Lemma. *Every partially ordered set $(P,\leq)$ contains at least one maximal chain.*

We refer to textbooks on set theory for more details and consequences of the above statement. Here we explain how it implies the general case of Theorem IV.3.1.

Proof. (of Theorem IV.3.1) Take the partially ordered set $(P,\leq)$ with

$$P = \{I: \ I \text{ is an ideal of } R, \text{ and } 1 \notin I\}.$$

As ordering on P we use the *inclusion relation*, so

$$I \leq J \text{ in } P \overset{\text{def}}{\Longleftrightarrow} I \subseteq J.$$

Observe that in the proof of the special case of IV.3.1 presented above, the set $\{I_n\}_{n\in\mathbb{N}}$ is actually a chain in P.

In the general case we use a similar idea: By Zorn's Lemma we even have a maximal chain K in P, say

$$K = \{I_n\}_{n\in X},$$

where the index set X could in principle be arbitrarily large. The chain K is a collection of ideals such that $\forall I, J \in K\colon\ I \subset J \vee J \subset I$. Incidentally, the partially ordered set P is not empty, since $\{0\} \subset R$ is an ideal. Hence the maximal chain K is nonempty as well. We now consider $M = \bigcup_{n\in X} I_n$.

We claim that M is the desired maximal ideal. To this end one first verifies that indeed M is an ideal, and this is done (check the details if necessary!) exactly as in the special case above. If the ideal M is not maximal then either $M = R$ or an ideal N exists with $M \subsetneq N \subsetneq R$. Again, checking that neither of these situations occurs is done quite analogous to the argument of the special case above: if $M = R$ then $1 \in M$ and therefore $1 \in I_n$ for some $n \in X$, contradicting the definition of P. The existence of an ideal N strictly between M and R would contradict the maximality of the chain K. We conclude that indeed M is maximal. ∎

IV.3.4 Corollary. *If R is a commutative ring (with 1) and $I \subset R$ is an ideal such that $I \neq R$, then R has a maximal ideal M with $I \subset M$.*

Proof. Applying IV.3.1, the ring R/I has a maximal ideal, which by II.3.10 has the form M/I with M some ideal of R such that $M \supset I$. Using II.3.10 we have that $R/M \cong (R/I)/(M/I)$ which is a field. Therefore M is maximal in R (Theorem IV.2.3). This proves IV.3.4.
(Different proof: apply Zorn's Lemma to the set of ideals $\neq R$ of R containing I.) ∎

IV.3.5 Corollary. *If R is a commutative ring with 1, then*

$$\bigcup_M M = R - R^\times,$$

where the union is taken over all maximal ideals M of R.

Proof. $\subset$: If M is maximal, then $M \subset R - R^\times$ by II.4.4 and IV.2.1(**M1**). So $\cup_M M \subset R - R^\times$.
$\supset$: If $a \in R - R^\times$ then $Ra \subsetneq R$ is an ideal of R.
Hence by IV.3.4 a maximal ideal M of R exists with $Ra \subseteq M$. Therefore $a \in \cup_M M$.
This finishes the proof of IV.3.5. ∎

IV.3.6 Example. Let $R = C([0,1])$ be the ring of continuous functions $f : [0,1] \to \mathbb{R}$. For $x \in [0,1]$ put

$$M_x = \{f \in R :\ f(x) = 0\}.$$

This is the kernel of the surjective ring homomorphism

$$R \to \mathbb{R}, \qquad f \mapsto f(x), \quad \text{so} \quad R/M_x \cong \mathbb{R}$$

and therefore $M_x \subset R$ is maximal. By Exercise 18 on page 60 it turns out that every maximal ideal of R is of this form. We have

$$R - \bigcup_{x\in[0,1]} M_x \ = \ \{f \in R :\ \forall\, x \in [0,1] : f(x) \neq 0\}.$$

This is evidently the group of units $R^\times$ of R, in accordance with IV.3.5. ∎

The next result dealing with the solvability of a system of polynomial equations over a field, shows a typical applications of Theorem refexmi.

IV.3.7 Corollary. *Let K be a field, $n, t \in \mathbb{Z}_{>0}$, and $f_1, f_2, \ldots, f_t \in K[X_1, X_2, \ldots, X_n]$. The following two statements are equivalent.*

(i) No $g_1, g_2, ..., g_t \in K[X_1, X_2, ..., X_n]$ exist with $g_1 f_1 + g_2 f_2 + \ldots + g_t f_t = 1$.
(ii) There exists a field L with $K \hookrightarrow L$ and $x_1, x_2, \ldots, x_n \in L$ such that $f_1(x_1, x_2, \ldots, x_n) = f_2(x_1, x_2, \ldots, x_n) = \ldots = f_t(x_1, x_2, \ldots, x_n) = 0$.

Proof. (ii) ⇒ (i). Given $L, x_1, ..., x_n$ as in (ii), suppose $g_j \in K[X_1, \ldots, X_n]$ exist with

$$g_1 f_1 + ... + g_t f_t = 1.$$

Substituting $x_1, x_2, ..., x_n$ for $X_1, X_2, ..., X_n$ one obtains $0 = 1$, a contradiction.
(i) ⇒ (ii). Let $I \subset K[X_1, ..., X_n]$ be the ideal generated by $f_1, f_2, ..., f_t$. Then (i) states that $1 \notin I$, so $I \neq K[X_1, ..., X_n]$. Applying IV.3.1 a maximal ideal M of $K[X_1, ..., X_n]$ exists with $I \subseteq M$. Put $L = K[X_1, ..., X_n]/M$. By IV.2.3 this is a field. Composing the ring homomorphisms

$$K \hookrightarrow K[X_1, ..., X_n] \longrightarrow L = K[X_1, ..., X_n]/M$$

one obtains a ring homomorphism $K \to L$, and this is injective by II.4.6.

We therefore may regard K as a subfield of L. Take $x_i = X_i + M \in L$, for $1 \leq i \leq n$. Then

$$f_j(x_1, ..., x_n) = f_j(X_1, ..., X_n) + M = 0 + M$$

because $f_j(X_1, ..., X_n) = f_j \in I \subseteq M$ for $1 \leq j \leq t$. This shows IV.3.7. ∎

IV.3.8 Example. Take

$$K = \mathbb{R}, \qquad n = t = 1, \qquad f_1 = X^2 + 1 \in \mathbb{R}[X].$$

Considering the degree one concludes that no $g_1 \in \mathbb{R}[X]$ exists with $g_1 f_1 = 1$. Hence condition (i) in Theorem IV.3.7 is satisfied. By the theorem an 'extension field' L of $\mathbb{R}$ exists containing an element $x \in L$ with $x^2 + 1 = 0$. Indeed one may take $L = \mathbb{C}, x = i$. This example shows that it is not possible in all cases to take $L = K$. —∎

IV.3.9 Remark. One can show that every maximal ideal of $\mathbb{C}[X_1, \ldots, X_n]$ has the form

$$M = (X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n) \qquad (a_i \in \mathbb{C}).$$

Hence in this case the maximal ideals correspond to the points of $\mathbb{C}^n$ (evidently the ideal M corresponds to the point $(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$).

A consequence of Theorem IV.3.7 in the present case is therefore: the polynomials $f_1, \ldots f_k \in \mathbb{C}[X_1, \ldots, X_n]$ have no common zero in $\mathbb{C}^n$, if and only if a polynomial relation $g_1 f_1 + g_2 f_2 + \ldots + g_t f_t = 1$ exists between the f_i.

IV.4 Exercises

1. Let R be a domain. Prove: the ideal of $R[X,Y]$ generated by X and Y equals
$$\{f \in R[X,Y] : f(0,0) = 0\}$$
and this is a prime ideal of $R[X,Y]$.
2. Let K be a field, $n \in \mathbb{Z}_{>0}$, and $\alpha_1, \alpha_2, ..., \alpha_n \in K$. Prove: the ideal of $K[X_1, X_2, ..., X_n]$ generated by $X_1 - \alpha_1, X_2 - \alpha_2, ..., X_n - \alpha_n$ is maximal.
3. Show: $5\mathbb{Z}[i] \subset \mathbb{Z}[i]$ is not a prime ideal.
4. Let K be a field. Prove that the ideal of $K[X,Y,Z]$ generated by Y and Z is prime but not maximal.
5. Verify for each of the following ideals of $\mathbb{Z}[X]$ whether it is a prime ideal, and whether it is a maximal ideal:
$$(X,3); \quad (X^2-3); \quad (5,X^2+3).$$
6. Let $a,b \in \mathbb{R}$ and $M = (X-a, Y-b) \subset \mathbb{R}[X,Y]$. Show that $f \in M \Leftrightarrow f(a,b) = 0$ and prove that M is maximal.
7. Check for each of the following ideals of $\mathbb{Q}[X,Y]$ whether it is a prime ideal, and whether it is a maximal ideal:
$$(X^2+1); \quad (X-Y, Y^2+1); \quad (X^2+1, Y^2+1); \quad (X^2+1, Y^2-2).$$
8. Let R be a commutative ring with 1 and $I \subset R$ an ideal. Prove: I is a prime ideal of $R \Leftrightarrow$ a field K and a ring homomorphism $f : R \to K$ exist with $f(1) = 1$ and $I = \mathrm{Ker}(f)$.
9. Suppose R is a commutative ring with 1 and $I \subset R$ an ideal and $\phi : R \to R/I$ the natural map. Let $J \subset R$ be a prime ideal such that $I \subset J$.
Show that $\phi(J)$ is a prime ideal of R/I and vice versa every prime ideal of R/I is of this form. (Hint: combine IV.1.5 and II.3.10).
10. Now do Exercise 9 with everywhere 'prime ideal' replaced by 'maximal ideal'.
11. Let $f : R_1 \to R_2$ be a (unitary) ring homomorphism between commutative rings, let $I_2 \subset R_2$ be an ideal, and $I_1 = f^{-1}(I_2) \subset R_1$.
    (a) Show: I_1 is an ideal in R_1, and R_1/I_1 is isomorphic to a subring of R_2/I_2.
    (b) Show: if I_2 is prime in R_2 then I_1 is prime in R_1.
    (c) Show by means of an example that (b) can be false if in both occurrences 'prime' is replaced by 'maximal'.
12. Suppose R is a Boolean ring (see Exercise 33 on page 16) with 1.
    (a) Prove: R is a domain $\Leftrightarrow R$ is a field $\Leftrightarrow R \cong \mathbb{F}_2$.
    (b) Let $I \subset R$ be an ideal. Show: I is a prime ideal $\Leftrightarrow I$ is a maximal ideal $\Leftrightarrow R/I \cong \mathbb{F}_2$.
13. Let R be a commutative ring with 1 and let $I \subset R$ be an ideal with $I \neq R$. We will assume that *every* $x \in R$ with $x \notin I$ satisfies $x^2 - 1 \in I$.
    (a) Prove: $R/I \cong \mathbb{F}_2$ or $R/I \cong \mathbb{F}_3$.
    (b) Is I a prime ideal of R?
14. Let R be a commutative ring with 1 and suppose $I \subset R$ is an ideal of *finite* index in R.
Show: I is a prime ideal $\Leftrightarrow I$ is a maximal ideal.
15. Suppose R is a commutative ring with $1 \neq 0$ and *every* ideal $I \neq R$ is a prime ideal. Prove that R is a field.

16. Let R be a commutative ring with 1 and assume at $I \cap J \neq \{0\}$ for all pairs of ideals $I \neq \{0\},\ J \neq \{0\}$ of R.
Show that $\{a \in R : a \text{ is a zero divisor }\} \cup \{0\}$ is a prime ideal of R.
17. Let R be the ring with additive group $(\mathbb{Q}, +, 0)$ and multiplication $xy = 0$ for all $x, y \in R$. Show: R contains no ideal M satisfying both (**M1**) and (**M2**) in IV.2.1. Why is this not in contradiction with Theorem IV.3.1 ?
18. Put $R = C([0,1])$ and let $M_x \subset R$ for $x \in [0,1]$ be as in Example IV.3.6.
    (a) Suppose $I \subset R$ is an ideal satisfying: $\forall x \in [0,1] :\ I \not\subset M_x$.
    Prove: $\forall x \in [0,1] : \exists f_x \in I :\ f_x(x) \neq 0$.
    Choose functions $f_x \in I$ as above. Show that $x_1, x_2, \ldots, x_n \in [0,1]$ exist with $\forall x \in [0,1] : \sum_{i=1}^{n} f_{x_i}(x)^2 > 0$. (Hint: use compactness of $[0,1]$, which means that if $[0,1] = \cup_{i \in I} U_i$ with U_i open, then a *finite* subset $J \subset I$ exists with $[0,1] = \cup_{j \in J} U_j$.)
    Conclude: $I = R$.
    (b) Suppose $M \subset R$ is a maximal ideal. Prove: $\exists x \in [0,1] : M = M_x$. Moreover, show that x is uniquely determined by M.
19. Let R be the ring of polynomial functions on the circle: $R = \mathbb{R}[X,Y]/I$ with $I = (X^2 + Y^2 - 1)$. Put $x := X + I$, $y = Y + I \in R$.
    (a) For $a, b \in \mathbb{R}$, show that $(x - a, y - b)$ is a maximal ideal of R if and only if $a^2 + b^2 = 1$.
    (b) For which $b \in \mathbb{R}$ is $(y - b)$ a maximal ideal of R ?
20. Let R be a commutative ring with 1 Suppose $a \in R$ satisfies $\forall n \in \mathbb{Z}_{>0} : a^n \neq 0$. Prove that R contains a prime ideal I such that $a \notin I$. (Hint: apply Zorn's Lemma to the set of ideals of R not containing any power of a.)
21. The *radical* denoted $\sqrt{0}$ of a commutative ring R with 1 is defined by
$$\sqrt{0} = \left\{a \in R : \exists n \in \mathbb{Z}_{>0} : a^n = 0\right\}.$$
Show that $\sqrt{0}$ is an ideal of R. Show that $\sqrt{0} = \cap_I I$, where I runs over all prime ideals of R (Hint: use Exercise 20).
22. The *Jacobson-radical* $J(R)$ of a commutative ring R with 1 is defined by
$$J(R) = \{x \in R : \forall r \in R : 1 + rx \in R^\times\}.$$
    (a) Let $x \in J(R)$ and let $M \subset R$ be a maximal ideal. Define the ideal I of R by $I := M + xR$. Show that $I \neq R$ and conclude that $x \in M$.
    (b) Let M be a maximal ideal of R and let $x \in M$. Show that $1 + x \notin M$.
    (c) Prove that $J(R) = \cap_M M$, the intersection over all maximal ideals M of R.
    (d) Show that $J(R)$ is an ideal of R.
23. Let R be a commutative ring with 1 and let $S \subset R$ be a nonempty subset satisfying $0 \notin S$ and $\forall s, t \in S$: $st \in S$.
Show that a prime ideal I of R exists with $I \cap S = \varnothing$. (Hint: use the ring $S^{-1}R$ introduced in Exercise 28 on page 16, and apply Theorem IV.3.1 and Exercise 11(b) above). How is this related to Exercise 20 above?
24. A commutative ring R with 1 is called *local* if $R - R^\times$ is an ideal of R.
    (a) Prove: R is local $\Leftrightarrow R$ has exactly one maximal ideal.
    (b) Let R be local and suppose $x \in R$ satisfies $x^2 = x$. Show: $x = 0$ or $x = 1$.
25. Let R be a commutative ring with 1 and let $I \subset R$ be a prime ideal. Put $S = R - I$.
    (a) Show that for all $s, t \in S$ also $st \in S$.
    (b) Show that the ring $S^{-1}R$ as defined in Exercise 28 on page 16 is a local ring (see Exercise 24 above).

26. Put $R = \{a/b \in \mathbb{Q} : a, b \in \mathbb{Z}, b \not\equiv 0 \bmod 5\}$. Show that R is a local ring. What is the maximal ideal M of R? Prove that $R/M \cong \mathbb{F}_5$.
27. Let X be a set. A *filter* on X is a collection $\mathscr{F}$ of subsets of X satisfying:

    (i) $X \in \mathscr{F}$ and $\varnothing \notin \mathscr{F}$;
    (ii) $A, B \in \mathscr{F} \Rightarrow A \cap B \in \mathscr{F}$;
    (iii) if $A \subset B \subset X$ and $A \in \mathscr{F}$, then $B \in \mathscr{F}$.

    An *ultrafilter* is a filter $\mathscr{F}$ with the additional property:

    $$\forall A, B \subset X : \quad (A \cup B \in \mathscr{F} \Rightarrow A \in \mathscr{F} \vee B \in \mathscr{F}).$$

    Let R be the ring $P(X)$ considered in Exercise 34 on page 17.

    (a) Let $\mathscr{F}$ be a collection of subsets of X. Prove:

    $$\mathscr{F} \text{ is a filter on } X \iff \{A \subset X : X - A \in \mathscr{F}\} \text{ is an ideal } \neq R \text{ of } R.$$

    Moreover, prove that

    $$\mathscr{F} \text{ is an ultrafilter on } X \iff \{A \subset X : X - A \in \mathscr{F}\} \text{ is a maximal ideal of } R.$$

    (Hint: use Exercise 12(b) on page 59.)

    (b) An ultrafilter $\mathscr{F}$ on X is called *free* if $\forall x \in X : \{x\} \notin \mathscr{F}$.
    Prove: free ultrafilters on X exist if and only if X is infinite.
    (Hint: apply Theorem IV.3.4).
28. Let K_x (for $x \in X$) be a collection of fields indexed by a set X. Let $R = \prod_{x \in X} K_x$ which is, with componentwise addition and multiplication, a commutative ring with 1. For an ultrafilter (see Exercise 27 above) $\mathscr{F}$ on X we define $I_{\mathscr{F}} \subset R$ by

    $$(\alpha_x)_{x \in X} \in I_{\mathscr{F}} \iff \{x \in X : \alpha_x = 0\} \in \mathscr{F}.$$

    Show that $I_{\mathscr{F}}$ is a maximal ideal of R. Moreover, prove that all maximal ideals of R are of the form $I_{\mathscr{F}}$ for some ultrafilter $\mathscr{F}$ on X.

V DIVISION IN RINGS

In the ring of integers $\mathbb{Z}$ the theorem of unique prime factorisation holds: every positive integer has a unique factorisation into prime factors. In this chapter we examine to what extend this result can be generalised to other rings R.

Throughout this chapter we restrict ourselves to *domains* R (see I.2.13).

V.1 Irreducible elements

It is natural to start by considering which elements of the domain R can play the role of 'prime numbers'. If $p \in \mathbb{Z}$ is a prime number, then:

(i) if $p = ab$ for $a, b \in \mathbb{Z}$ then $a = \pm 1$ or $b = \pm 1$.
(ii) $\mathbb{Z}p := \{np : n \in \mathbb{Z}\} \subset \mathbb{Z}$ is a prime ideal.

Moreover, if $p \in \mathbb{Z}_{>0}$ satisfies either (i) or (ii), then p is prime (and hence p satisfies both (i) and (ii)). We now generalise the property (i), and we present examples where (i) holds but (ii) does not, see Example V.1.5 and Exercises 1 and 2 on page 75.

V.1.1 Definition. An element a of a domain R is called *irreducible* if a is not a unit, and for all $b, c \in R$ such that $bc = a$ either $b \in R^\times$ or $c \in R^\times$.

In other words: an element is irreducible if it only allows 'trivial' factorisations, such as $5 = (-1)\cdot(-5)$. The irreducible elements of $\mathbb{Z}$ are exactly the prime numbers p and their opposites $-p$.

V.1.2 Example. Let R be a domain and $f, g \in R[X]$. Then $\deg(fg) = \deg(f) + \deg(g)$ and $\deg(1) = 0$. So units in $R[X]$ are precisely the polynomials which have degree 0 and moreover have an inverse. In other words, the units in $R[X]$ are just the units in R, i.e.,

$$(R[X])^\times = R^\times.$$

A polynomial of degree 1 is *not* necessarily irreducible: $2X - 2 = 2\cdot(X-1)$ is reducible in $\mathbb{Z}[X]$, since the factors 2 and $X - 1$ are not units in this ring. However, $2X - 2$ is irreducible in $\mathbb{Q}[X]$ (here 2, and any polynomial of degree 0, is a unit).

More generally for $R = K$ a field, every polynomial of degree 0 is a unit, and as a consequence every polynomial of degree 1 is irreducible. For polynomials of higher degree it is in general a delicate matter to determine whether they are irreducible or not, see Section V.5 below. ∎

V.1.3 Theorem. *Let K be a field and $f \in K[X]$ such that* $\deg(f) = 2$ *or* $\deg(f) = 3$. *Then f is irreducible in $K[X]$ if and only if f has no zero in K.*

Proof. If $a \in K$ is a zero of f then $f = (X - a)g$ (see III.5.1) hence f is reducible.

Now assume f has no zero in K. From $\deg(f) > 0$ one concludes that f is not a unit. Suppose $f = gh$, where we will assume without loss of generality that $\deg(g) \leq \deg(h)$. If $\deg(g) \neq 0$ then $\deg(g) = 1$ since $\deg(f) = \deg(g) + \deg(h)$. Hence, g has a zero in K, and therefore so does f, contradicting the assumption. It follows that $\deg(g) = 0$ so g is a unit in $K[X]$. We conclude that f is irreducible. ∎

The next result relates irreducible elements to prime ideals. As the example following it shows, the reverse statement does *not* hold in general. In fact the remainder of this chapter discusses rings in which the converse of the following theorem indeed holds.

V.1.4 Theorem. *Let $a \in R$ such that Ra is a prime ideal $\neq (0)$. Then a is irreducible.*

Proof. We have $a \neq 0$. Applying property (**P1**) of IV.1.1 to the prime ideal Ra shows $Ra \neq R$, hence a is not a unit (see II.4.4). Now assume $bc = a$ for certain $b, c \in R$. Since $a \in Ra$, property (**P2**) implies $b \in Ra$ or $c \in Ra$. Interchanging b and c if necessary, we may assume $b \in Ra$. This means $b = ra$ for some $r \in R$. The conditions $bc = a$ and $b = ra$ imply (since R is commutative) that $(rc - 1)a = 0$. Hence because R is a domain and $a \neq 0$ one concludes $rc = 1$. This shows that c is a unit.

So, in every factorisation $bc = a$ either $b \in R^\times$ or $c \in R^\times$, which implies that a is irreducible. This proves Theorem V.1.4. ∎

V.1.5 Example. Consider the ring

$$R = \left\{ \sum_{i=0}^{n} a_i X^i \in \mathbb{Q}[X] : \ a_1 = 0, \ n \in \mathbb{Z}_{\geq 0} \right\}$$

(incidentally, instead of $\mathbb{Q}$ one may take an arbitrary field K here); it is not hard to verify that indeed this defines a subring of $\mathbb{Q}[X]$. Observe that $X \notin R$.

We claim that X^2 is an irreducible element of R, however RX^2 is *not* a prime ideal in R.

First a proof that X^2 is irreducible: check for yourself that X^2 is not a unit. It remains to show: if

$$X^2 = f \cdot g, \quad \text{with} \quad f, g \in R, \quad \text{then} \quad f \in R^\times \quad \text{or} \quad g \in R^\times.$$

From $X^2 = f \cdot g$ it follows that $\deg(f) + \deg(g) = 2$. However, R contains no polynomials of degree 1, and therefore either f or g has degree 0, let us say this holds for f. Then $f \in \mathbb{Q}$ and $f \neq 0$, hence f has an inverse in $\mathbb{Q}$ so certainly in R. This shows $f \in R^\times$ and as a consequence X^2 is irreducible in R. (Evidently X^2 is not irreducible in $\mathbb{Q}[X]$, since $X^2 = X \cdot X$.)

Now we show that RX^2 is not a prime ideal of R. Note that $X^3 \cdot X^3 \in R \cdot X^2$ (because $X^4 \in R$). Were RX^2 a prime ideal of R, then property (**P2**) of IV.1.1 (with $a = b = X^3$) would imply $X^3 \in RX^2$, so $X \in R$, a contradiction. ∎

V.2 Principal ideal domains

This section discusses an important class of rings in which every irreducible element generates a prime ideal, and in fact even maximal ideal.

V.2.1 Definition. A *principal ideal domain* (PID) is a domain R in which every ideal is principal (see Definition II.2.8).

V.2.2 Remark. In some textbooks the term “principal ideal ring" is used for domains in which every ideal is principal. Other texts allow a principal ideal ring to have zero divisors.

V.2.3 Example. In Theorem II.4.2 we saw that $\mathbb{Z}$ is a principal ideal domain. And Example II.4.3 shows that $\mathbb{R}[X,Y]$ is not a principal ideal domain. Every field is a principal ideal domain, as follows trivially from II.4.5. By III.4.1 $K[X]$ is a principal ideal domain for every field K. ▬

The next result shows that in a principal ideal domain several of the notions introduced here coincide.

V.2.4 Theorem. *Let R be a principal ideal domain, and $a \in R, a \neq 0$. The next three statements are equivalent:*

*(*i*) Ra is a maximal ideal of R;*
*(*ii*) Ra is a prime ideal of R;*
*(*iii*) a is irreducible in R.*

Proof. (i)⇒ (ii): this is an immediate consequence of IV.2.7.
(ii)⇒(iii): this is exactly Theorem V.1.4.

So far we did not use the condition that R is a *principal ideal domain*. This will be used to prove the remaining implication (iii)⇒(i):
Given is that $a \in R$ is irreducible. We must show that the ideal Ra satisfies the conditions (**M1**) and (**M2**) of IV.2.1.
(**M1**) a is irreducible, hence it is not a unit. As a consequence $Ra \neq R$.
(**M2**) Suppose J is an ideal of R with $Ra \subset J \subset R$. We claim that $J = Ra$ or $J = R$.

As R is a principal ideal domain, $J = Rb$ for some $b \in R$. Then $a \in Ra \subset J = Rb$ implies $a \in Rb$, hence $a = rb$ for some $r \in R$. The irreducibility of a then shows that either $r \in R^\times$ or $b \in R^\times$.

In case $r \in R^\times$ we have $b = r^{-1}a \in Ra$ and therefore $J = Rb \subset Ra$, implying $J = Ra$. In case $b \in R^\times$ we have $J = Rb = R$.
This verifies (**M2**) and completes the proof. ■

In particular one concludes that in principal ideal domains the converse of IV.2.7 holds for ideals $\neq \{0\}$:

V.2.5 Corollary. *In a principal ideal domain every prime ideal $\neq \{0\}$ is maximal.*

Proof. Noting that every ideal in such a ring is principal, this follows from the implication (ii)⇒(i) in V.2.4. ■

V.2.6 Example. Theorem V.2.4 shows that the ring R from Example V.1.5 is not a principal ideal domain (note that R is a domain). In fact, the ideal generated by X^2 and X^3 is not principal, see Exercise 4 on page 75. ▬

An important application of Theorem V.2.4 is the construction of a field in which a given polynomial has a zero.

V.2.7 Theorem. *Let K be a field and $f \in K[X]$ an irreducible polynomial. Put*

$$\alpha := X + (f) \in K[X]/(f).$$

Then $L := K[X]/(f)$ is a field and $K \subset L$ is a subring and α is a zero of f in L.

Proof. Since $K[X]$ is a principal ideal domain and f is irreducible, the ideal (f) is maximal and therefore $L = K[X]/(f)$ is a field. The inclusion $K \subset L$ is given by $a \mapsto a + (f)$ $(a \in K)$; one simply writes a instead of $a + (f)$ or $\bar{a}$.

Using $K \subset L$ one considers f as an element of $L[X]$. In L it holds that

$$\alpha^i = (X + (f))^i := X^i + (f) \qquad \text{and} \qquad a_i(X^i + (f)) = a_i X^i + (f) \qquad (a_i \in K).$$

Writing $f = a_0 + a_1 X + \ldots + a_n X^n$ it follows that

$$f(\alpha) = a_0 + a_1 X + \ldots + a_n X^n + (f) = f + (f) = 0 + (f),$$

which shows that $\alpha \in L$ is a zero of f. This finishes the proof. ∎

V.2.8 Example. Put $K = \mathbb{R}$, $f = X^2 + 1 \in \mathbb{R}[X]$, $L = \mathbb{R}[X]/(X^2+1)$, and $\alpha := X + (X^2+1)$. Then $\mathbb{R}$ can be considered as a subring of L. Every element of L has a unique representation $a + bX + (X^2+1) = a + b\alpha$. Then

$$\begin{aligned} \alpha^2 = (X + (X^2+1))^2 \quad &:= X^2 + (X^2+1) \\ &= -1 + 1 \cdot (X^2+1) + (X^2+1) \\ &= -1 + (X^2+1), \end{aligned}$$

and therefore $\alpha^2 = -1 \in L$. This shows that α is indeed a zero of f. Note that we already showed $L \cong \mathbb{C}$ (see III.4.3); the given isomorphism sends $\alpha = X + (f)$ to $i \in \mathbb{C}$ and i is a zero of $X^2 + 1$ in $\mathbb{C}$.

More generally, compare Exercise 12 on page 49, if $g = X^2 + bX + c \in \mathbb{R}[X]$ has no zero in $\mathbb{R}$ then $\mathbb{R}[X]/(X^2 + bX + c) \cong \mathbb{C}$ and $X + (X^2 + bX + c)$ corresponds to a zero $z \in \mathbb{C}$ of g. ▬

V.2.9 Example. Let $\mathbb{F}_p$ be the field $\mathbb{Z}/p\mathbb{Z}$, for p a prime number (see I.2.11). If $p > 2$, then $a \in \mathbb{F}_p$ exists such that $X^2 - a \in \mathbb{F}_p[X]$ is irreducible (which in this case means $X^2 - a$ has no zero in $\mathbb{F}_p$). Indeed, the set $\{x^2 : x \in \mathbb{F}_p\}$ contains at most $1 + \frac{p-1}{2}$ (pairwise distinct) elements, since $x^2 = (-x)^2$ (in fact the set consists of precisely $1 + (p-1)/2$ elements). Hence some $a \in \mathbb{F}_p$ exists which is not a square, and for such a the polynomial $X^2 - a$ has no zero in $\mathbb{F}_p$. For $p = 3, 5, 7$ one can take $a = 2, 2, 3$, respectively.

As a result, for every prime $p > 2$ a field consisting of exactly p^2 elements exists, namely $\mathbb{F}_p[X]/(X^2 - a)$ with a as above. Every element r in this field can be written in a unique way as $r = c + b\alpha$ for some $c, b \in \mathbb{F}_p$ and $\alpha := X + (X^2 - a)$. For $p = 2$ we already saw a field consisting of p^2 elements, see III.3.6. We will see later, IX.1.1, that up to isomorphism only one field containing exactly p^2 elements exists. ▬

V.3 Unique factorization domains

We will present a general method for showing that in certain rings irreducible elements generate prime ideals.

V.3.1 Definition. A *unique factorization domain* (or simply: *factorization domain*) is a domain R with the property that every $a \in R, a \neq 0$, can be written as a product of a unit and a finite number of irreducible elements:

$$a = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t, \qquad u \in R^\times,\ t \in \mathbb{Z}_{\geq 0},\ p_i \in R \text{ irreducible}$$

and moreover such a factorization is assumed to be unique up to its order and up to units, i.e., if

$$a = v \cdot q_1 \cdot q_2 \cdot \ldots \cdot q_s, \qquad v \in R^\times,\ s \in \mathbb{Z}_{\geq 0},\ q_i \in R \text{ irreducible},$$

is another factorization, then $s = t$ and a permutation σ of $\{1,2,\ldots,t\}$ exists such that

$$p_i = v_i \cdot q_{\sigma(i)} \qquad \text{for certain units } v_i \in R^\times,\ i = 1,2,\ldots,t.$$

(and as a consequence $v = uv_1v_2\cdots v_t$.)

The given factorization of a is called the *prime factorization* of a, in analogy with the prime factorization in $\mathbb{Z}$. (Indeed the ideals Rp_i are prime ideals, see Theorem V.3.2.)

Roughly speaking: a unique factorization domain is a domain in which the unique factorization into primes holds. Note that the appropriate care with respect to *units* used in the definition is not needed in the case $R = \mathbb{Z}$, since there we restricted to *positive* integers. However, a similar restriction is not available in more general domains.

V.3.2 Theorem. *Let R be a unique factorization domain, and $a \in R$. Then*

$$a \text{ is irreducible} \iff Ra \text{ is a prime ideal } \neq (0).$$

Proof. $\Leftarrow$: this holds in general, see V.1.4.
$\Rightarrow$: Suppose $a \in R$ is irreducible. Then $a \neq 0$, and it remains to show that Ra is a prime ideal of R. Evidently (**P1**): $Ra \neq R$ holds, since a is not a unit.

We verify (**P2**). Suppose $b,c \in R$ satisfy $bc \in Ra$. We must show that $b \in Ra$ or $c \in Ra$. This clearly holds if $b = 0$ or $c = 0$, so we may assume $b,c \neq 0$. Then also $bc \neq 0$, hence $bc \in Ra$ can be written as $bc = da$ with $d \in R$, $d \neq 0$. Factorizing d into irreducible elements (and a unit), it follows that bc has a factorization into irreducible elements including the element a. Any factorization of bc into irreducible factors (and a unit) is obtained by combining one such factorization for b and one for c. The unicity of the factorization therefore implies that the element a (multiplied by a unit if necessary) occurs here, so a appears in the factorization of either b or c. This shows $b \in Ra$ or $c \in Ra$, proving V.3.2. ∎

The conclusion is that in a unique factorization domain the converse of Theorem V.1.4 holds. To verify that certain domains are indeed unique factorization domains, the next lemma will be used.

V.3.3 Lemma. *Let R be a domain in which every $a \in R$, $a \neq 0$ can be written as a product of a unit and a finite number of elements:*

$$a = u \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_t, \quad u \in R^\times,\ t \in \mathbb{Z}_{\geq 0},\ p_i \in R$$

where moreover for all $i = 1, 2, \ldots, t$ it holds that p_iR is a prime ideal.

Then R is a unique factorization domain.

Proof. Note that in a factorization as given, $a \neq 0$ implies that all $p_i \neq 0$. Hence the prime ideals p_iR are nonzero, so Theorem V.1.4 shows that the p_i are irreducible. It remains to show that the given factorization is *unique*. Suppose $a = up_1\cdots p_t$ has another factorization:

$$up_1p_2\cdots p_t = vq_1q_2\cdots q_s$$

with $u,v \in R^\times$, $t,s \in \mathbb{Z}_{\geq 0}$, p_i irreducible and Rp_i a prime ideal for $(1 \leq i \leq t)$, and irreducible q_j $(1 \leq j \leq s)$.

We will show that $s = t$, and that the q_j's up to units coincide with the p_i's. We use induction w.r.t. t.

If $t = 0$ then $vq_1q_2...q_s = u$ is a unit. Since irreducible elements are not units, this is only possible when $s = 0$ and $v = u$, as desired.

Now let $t > 0$. Then $q_1q_2 \cdot \ldots \cdot q_s = v^{-1} \cdot up_1p_2 \cdot \ldots \cdot p_t \in Rp_t$ which is a prime ideal. If $s = 0$ then this implies Rp_t contains a unit, contradicting (**P1**) of IV.1.1). Hence $s > 0$. By (**P2**) of IV.1.1 the product $q_1 \cdot q_2 \cdot \ldots \cdot q_s$ being in Rp_t implies that one of the factors, say q_s, is in Rp_t: $q_s = r \cdot p_t$ for some $r \in R$. Since q_s is irreducible and p_t is not a unit, r must be a unit. Substituting $r \cdot p_t$ for q_s and using that R is a domain, we conclude

$$up_1p_2 \cdots p_{t-1} = (rv)q_1q_2 \cdots q_{s-1}, \qquad rv \in R^\times.$$

From the induction hypothesis therefore $t - 1 = s - 1$ and the $p_1, p_2, ..., p_{t-1}$ coincide up to units with the $q_1, q_2, ..., q_{s-1}$. As a consequence $s = t$ and $p_1, p_2, ..., p_t$ coincides up to units with $q_1, q_2, ..., q_s$. This proves Lemma V.3.3. ∎

The next result uses this lemma:

V.3.4 Theorem. *Every principal ideal domain is a unique factorization domain.*

Proof. Let R be a principal ideal domain. It suffices to show that every $r \in R$ with $r \neq 0$ has a factorization $r = up_1 \cdot \ldots \cdot p_t$ (with Rp_i prime ideals): unicity then follows from Lemma V.3.3.

Assume $a_1 \in R$, $a_1 \neq 0$ does not have a factorization as desired. Then the ideal Ra_1 is not equal to R (otherwise a_1 would be a unit and that yields the desired factorization). Applying Corollary IV.3.4 a maximal ideal M exists with $Ra_1 \subset M$. Since R is a principal ideal domain, $M = Rp_1$ for some $p_1 \in R$. As $a_1 \in Ra_1 \subset M = Rp_1$ we can write $a_1 = a_2p_1$ with $a_2 \in R$. Now $Ra_1 \subset Ra_2$ and using that p_1 is not a unit, $Ra_1 \neq Ra_2$.

We repeat the argument above: $a_2 \neq 0$ and a_2 is not a unit (recall that by assumption a_1 does not admit a factorization as product of a unit times a finite set of generators of prime ideals), so a maximal ideal $Rp_2 \supset Ra_2$ exists, et cetera. Continuing in this way one obtains a chain of ideals $(Ra_n)_{n=1}^\infty$ with $Ra_n \subset Ra_{n+1}$ but $Ra_n \neq Ra_{n+1}$ (for all n). Put

$$I := \bigcup_{n \geq 1} Ra_n \qquad (\subset R).$$

It is easy to verify (check for yourself!) that I is an ideal of R. Again using that R is a principal ideal domain, $d \in R$ exists with

$$I = Rd.$$

Now I is the union of all Ra_n, hence m exists with $d \in Ra_m$. But then

$$Ra_m \subset Ra_{m+1} \subset I = Rd \subset Ra_m,$$

contradicting $Ra_m \neq Ra_{m+1}$.

The assumption that some element $a_1 \neq 0$ in R has no factorization as a product of a unit times a finite number of generators of prime ideals therefore leads to a contradiction. We conclude that every $r \in R$, $r \neq 0$, admits such a factorization, which proves the result. ∎

Let K be a field. Then by III.4.1 $K[X]$ is a principal ideal domain, hence by Theorem V.3.4 $K[X]$ is a unique factorization domain. Any irreducible $g \in K[X]$ has positive degree. The leading coefficient a_n of such a g is a unit, so g can be written in a unique way as $g = a_nh$ with h monic and irreducible. The prime factorization of any nonzero $f \in K[X]$ can therefore be given as

$$f = uh_1^{n_1}h_2^{n_2} \ldots h_k^{n_k},$$

with $u \in K^\times = K[X]^\times$ and the h_i pairwise distinct monic irreducible polynomials. Moreover this factorization is unique (up to permuting the h_i's). This is similar to the situation in $\mathbb{Z}$ where one takes *positive* irreducible elements after multiplying with an appropriate unit ± 1.

V.3.5 Theorem. *Let K be a field and $f \in K[X]$ non-constant. Suppose $f = uh_1^{n_1} \dots h_k^{n_k}$ is the prime factorization of f (so u is a unit and the h_j are irreducible and pairwise distinct).*

Then $k \geq 1$ and

$$K[X]/(f) \cong \big(K[X]/(h_1^{n_1})\big) \times \dots \times \big(K[X]/(h_k^{n_k})\big).$$

Proof. We use induction w.r.t. the number k of irreducible factors of f. Since f is non-constant and all constants in $K[X]$ have degree ≤ 0, we have $k \geq 1$. For $k = 1$ the theorem trivially holds.

Now let $k > 1$ and write

$$f = (uh_1^{n_1} \dots h_{k-1}^{n_{k-1}})h_k^{n_k} =: f_{k-1}h_k^{n_k}.$$

Define the ideals I, J in $K[X]$ by

$$I = (f_{k-1}) \qquad \text{and} \qquad J = (h_k^{n_k}).$$

We claim that $I + J = K[X]$, which will allow us to apply the Chinese remainder theorem II.4.12 to $K[X]/(f) = K[X]/IJ$.

Since $K[X]$ is a principal ideal domain,

$$I + J = (g)$$

for some $g \in K[X]$. We have $h_k^{n_k} \in J \subset (g)$, so $r \in K[X]$ exists with $h_k^{n_k} = rg$. Considering the (unique) prime factorization of r and of g in $K[X]$ and using that h_k is irreducible, it follows that

$$g = vh_k^m$$

for some $m \leq n_k$ and some unit v.

On the other hand also $f_{k-1} \in I \subset (g)$, so $s \in K[X]$ exists with $f_{k-1} = sg$, which means

$$uh_1^{n_1} \dots h_{k-1}^{n-1} = svh_k^m.$$

Here the h_i are monic and irreducible and $K[X]$ is a unique factorization domain, hence $m = 0$. This shows that g is a unit in $K[X]$, so indeed $I + J = (g) = K[X]$.

The Chinese remainder theorem II.4.12 now implies

$$K[X]/(f) = K[X]/(f_{k-1}h_k^{n_k}) \cong K[X]/(f_{k-1}) \times K[X]/(h_k^{n_k}).$$

Using the induction hypothesis for the ring $K[X]/(f_{k-1})$ now yields

$$K[X]/(f) \cong K[X]/(h_1^{n_1}) \times K[X]/(h_2^{n_2}) \dots \times K[X]/(h_k^{n_k}),$$

as desired. This finishes the proof of Theorem V.3.5. ∎

V.4 Polynomials over unique factorization domains

Given any field K, Theorem III.4.1 states that $K[X]$ is a principal ideal domain and therefore by Theorem V.3.4 we have that $K[X]$ is a unique factorization domain. The current section intends to prove a more general result.

V.4.1 Theorem. *If R is a unique factorization domain, so is $R[X]$.*

V.4.2 Corollary. *For every integer $n > 0$ and every unique factorization domain R the ring $R[X_1, X_2, \dots, X_n]$ is a unique factorization domain. In particular the rings $\mathbb{Z}[X_1, X_2, \dots, X_n]$ and $K[X_1, X_2, \dots, X_n]$ (K a field) are unique factorization domains.*

Proof. (of Corollary V.4.2.) This is immediate from Theorem V.4.1 using induction w.r.t. n. ∎

We will prove Theorem V.4.1 using the following strategy. Given the unique factorization domain R, let $K := Q(R)$ be its field of fractions as introduced in I.3. Since K is a field we know that $K[X]$ is a unique factorization domain. Hence given a nonzero $f \in R[X]$, there exists a factorization of f into irreducible elements of $K[X]$. What remains, is to modify this factorization (by clearing the denominators of the coefficients of the factors we have in $K[X]$), and in this way obtain a factorization in $R[X]$.

V.4.3 Definition. Elements a, b in a unique factorization domain R are called *associated* if $a = ub$ for some $u \in R^\times$.

In the remainder of this section we fix a unique factorization domain R. It is not difficult to see that 'being associated' defines an equivalence relation on R. Let $P \subset R$ be a set of irreducible elements of R; moreover we take P in such a way that every irreducible element of R is associated with precisely one element of P.

V.4.4 Example. In case $R = \mathbb{Z}$ one can take P the set of all *positive* irreducible elements, in other words the set of all prime numbers.

In case $R = K[X]$ and K a field, the set of all monic irreducible polynomials is such a P. ∎

V.4.5 Definition. For R and P as above, all nonzero $a, b \in R$ have a unique factorization

$$a = u \cdot \prod_{p \in P} p^{n(p)}, \qquad b = v \cdot \prod_{p \in P} p^{m(p)}$$

with $n(p), m(p) \in \mathbb{Z}_{\geq 0}$, only finitely many $n(p), m(p)$ different from 0, and $u, v \in R^\times$. The *greatest common divisor* (gcd) of a and b is defined as

$$\gcd(a, b) := \prod_{p \in P} p^{\min\{n(p), m(p)\}} \qquad (\in R).$$

For a different choice P' instead of P, the $p \in P$ will be changed into elements $p' \in P'$ which are associated with the original ones. Then $n(p) = n(p')$ if p, p' are associated, and the gcd will change by at most a unit. Exercise 7 on page 76 explains the terminology 'gcd'.

V.4.6 Definition. With R and P as above, let $f = \sum_{i=0}^{n} a_i X^i \in R[X]$, $f \neq 0$ and take d the greatest common divisor of the coefficients $a_0, a_1, \ldots, a_n$. This d is called the *content* of f, notation:

$$\mathrm{co}(f) := \gcd(a_0, a_1, \ldots, a_n).$$

A polynomial with content a unit is called *primitive*.

Note that the content $\mathrm{co}(f)$ of a nonzero polynomial f depends on the choice of $P \subset R$, hence it is only defined up to multiplication by a unit in R. Any nonzero polynomial f can be written as $f = rg$ with $r = \mathrm{co}(f) \in R$ and $g \in R[X]$ a primitive polynomial. The next lemma extends this observation to polynomials with coefficients in K.

V.4.7 Lemma. *Any $f \neq 0$ in $K[X]$ can be written as*

$$f = d \cdot f_0, \qquad \textit{with} \quad d \in K^\times \quad \textit{and} \quad f_0 \in R[X] \ \textit{primitive.}$$

Up to multiplication by units in R this way of writing f is unique.

Proof. Let c be the product of the denominators of the coefficients of f. Then

$$cf \in R[X], \qquad \text{and} \quad cf = \mathrm{co}(cf) \cdot f_0 \qquad \text{with} \quad f_0 \in R[X]$$

a primitive polynomial. Hence in $K[X]$ one finds $f = c^{-1} \cdot cf = (c^{-1} \cdot \mathrm{co}(cf))f_0$, so taking $d = c^{-1} \cdot \mathrm{co}(cf)$ yields an expression as desired.

Now suppose $d \cdot f_0 = e \cdot g_0$, with $d, e \in K^\times$ and $f_0, g_0 \in R[X]$ primitive. We claim that $d = e \cdot u$, $f_0 = u^{-1} \cdot g_0$ for some $u \in R^\times$, proving the asserted uniqueness. Multiplying both d and e by a suitable nonzero element of R we may assume $d, e \in R$. Then d and e both equal (up to a unit in R) the content of $d \cdot f_0 = e \cdot g_0$, hence they are associated in R. This proves Lemma V.4.7. ∎

The next lemma states how the content of a product of polynomials is related to the product of the contents.

V.4.8 Lemma. *If $f, g \in R[X]$ are nonzero polynomials, then* $\mathrm{co}(fg)$ *and* $\mathrm{co}(f) \cdot \mathrm{co}(g)$ *are associated in R. In particular, the product of primitive polynomials is primitive as well.*

Proof. Suppose $f = \sum a_i X^i$ and $g = \sum b_j X^j$ are primitive but $f \cdot g = \sum c_k X^k$ is not. Then an irreducible element $p \in R$ exists dividing all coefficients c_k of $f \cdot g$. So $c_k \in Rp$ for all k. Let $\overline{f} = \sum \overline{a}_i X^i \in (R/pR)[X]$ and $\overline{g} = \sum \overline{b}_j X^j \in (R/pR)[X]$ (here $\overline{a} = (a \bmod pR) \in R/pR$, for $a \in R$). In $(R/pR)[X]$ we now have

$$\overline{f} \cdot \overline{g} = (\sum \overline{a}_i X^i) \cdot (\sum \overline{b}_j X^j) = \sum \overline{c}_k X^k = \sum \overline{0} \cdot X^k = \overline{0}.$$

By Theorem V.3.2 Rp is a prime ideal of R, hence (see Theorem IV.1.5) R/pR is an integral domain. Therefore $(R/pR)[X]$ is a domain as well. Hence the product $\overline{f} \cdot \overline{g}$ can only be zero if one of the factors $\overline{f}$ or $\overline{g}$ is zero; say $\overline{f}$. Then all $\overline{a}_i$ equal $\overline{0}$, which means all a_i are divisible by p. This contradicts the assumption that f is primitive. We conclude that the product of primitive polynomials is primitive as well.

If f, g are arbitrary nonzero polynomials then

$$f \cdot g = \mathrm{co}(f) f_0 \cdot \mathrm{co}(g) g_0 = (\mathrm{co}(f) \cdot \mathrm{co}(g)) f_0 \cdot g_0.$$

Here f_0 and g_0 are primitive, hence so is $f_0 g_0$. As a consequence, up to multiplication by a unit the content of $f \cdot g$ equals $\mathrm{co}(f) \cdot \mathrm{co}(g)$. This finishes the proof of Lemma V.4.8. ∎

V.4.9 Lemma. *Any nonzero $f \in R[X]$ can be written as*

$$f = u \cdot p_1 p_2 \cdots p_s \cdot g_1 g_2 \cdots g_t$$

with $u \in R^\times, s, t \in \mathbb{Z}_{\geq 0}$ and $p_1, p_2, ..., p_s$ irreducible elements of R and $g_1, g_2, ..., g_t$ primitive polynomials in $R[X]$ which are irreducible in $K[X]$.

Moreover, this way of writing f is unique up to ordering and multiplication by units of R.

Proof. Since $K[X]$ is a unique factorization domain, f can be written as

$$f = d \cdot g_1 g_2 \cdots g_t,$$

with $d \in K[X]^\times = K^\times, t \in \mathbb{Z}_{\geq 0}$, and $g_1, g_2, ..., g_t \in K[X]$ irreducible. Up to ordering the g_i's and multiplying by elements of $K^\times$ this factorization is unique. Writing every g_i in the form as given by Lemma V.4.7, one concludes that we can assume the g_i are primitive in $R[X]$ (changing d if necessary). Moreover with this additional

constraint the g_i are fixed up to units in R. Since $f = dg_1g_2\cdots g_t$, the same holds for d.

As each g_i is primitive, so is (by V.4.8) $g_1g_2\cdots g_t$. Hence $f = d\cdot(g_1g_2\cdots g_t)$ is the unique way of writing f as given in V.4.7. Hence d equals the content of f; in particular $d \in R$. Now factor d in R as

$$d = u\cdot p_1p_2\cdots p_s \qquad (u \in R^\times,\ s \in \mathbb{Z}_{\geq 0},\ p_i \in R \text{ irreducible})$$

(this is again unique up to ordering and units, as R is a unique factorization domain). Now $f = up_1p_2\cdots p_sg_1g_2\cdots g_t$ is the required (unique) factorization. This shows V.4.9. ■

Using the lemmas above we now show the main result of this section.

Proof. (of Theorem V.4.1.) Let $f \in R[X], f \neq 0$. We claim that the factorization of f as given by Lemma V.4.9 is in fact a factorization into irreducible elements of $R[X]$. For this, we only have to show that the irreducible elements of $R[X]$ are precisely the irreducible elements p of R and the primitive polynomials $g \in R[X]$ which are irreducible in $K[X]$.

Assume $f \in R[X]$ is irreducible, and write f as in V.4.9. Then $s+t \neq 0$ (since f is not a unit), and $s+t < 2$ (otherwise we obtain a factorization of f into two non-units). So $s+t = 1$, which shows that f equals (up to a unit) some p or some g as described above. So indeed the irreducible elements of $R[X]$ have the desired form.

Vice versa, let p (respectively g) be an irreducible element of R (respectively a primitive, in $K[X]$ irreducible polynomial from $R[X]$). This is not a unit in $R[X]$ since $R[X]^\times = R^\times$. If it can be written as a product f_1f_2 of two non-units in $R[X]$, one immediately arrives at a contradiction with the uniqueness given in V.4.9 by combining the factorizations of f_1 and f_2 into one for p (or g) $= f_1f_2$. We conclude that p (respectively g) is irreducible in $R[X]$. This shows V.4.1. ■

We draw some important conclusions from the proof of Theorem V.4.1.

V.4.10 Corollary. *Let R be a unique factorization domain with field of fractions K, and suppose $f \in R[X]$ is* primitive. *Then*

f is irreducible in $K[X] \iff f$ is irreducible in $R[X]$.

Proof. $\Leftarrow$: We saw earlier that any irreducible $f \in R[X]$ is either irreducible in $K[X]$ or it is an irreducible element of R. In the present case the latter possibility does not occur since f is primitive.

$\Rightarrow$: Suppose $f = g\cdot h$ with $g,h \in R[X]$. Since f is irreducible in $K[X]$ one of these factors, say g, is a unit in $K[X]$. So $g \in K^\times \cap R[X] = R - \{0\}$. From $f = g\cdot h$ one concludes that g divides the content of f. However $\mathrm{co}(f) = 1$, so $g \in R^\times$. Hence f is irreducible in $R[X]$. This proves Corollary V.4.10. ■

V.4.11 Corollary. (Lemma of Gauss) *Let R be a unique factorization domain with field of fractions K, and suppose $f \in R[X]$ is monic.*
If $g \in K[X]$ is monic and g divides f, then $g \in R[X]$.

Proof. Since $g|f$ and both polynomials are monic, a monic $h \in K[X]$ exists with $f = gh$. By V.4.7 $u,v \in K^\times$ exist with $u\cdot g$ and $v\cdot h$ primitive in $R[X]$. These polynomials have leading coefficients u and v, hence $u,v \in R$. Now on the one hand $f \in R[X]$ is monic, hence primitive. On the other hand, using Lemma V.4.8 also $uv\cdot f = (ug)\cdot(vh)$ is primitive. This is only possible if $uv \in R^\times$, which implies that u and v are units in R. We conclude: $g = u^{-1}\cdot ug \in R[X]$. This shows V.4.11. ■

V.5 Factorizing and irreducibility of polynomials

We discuss some practical methods to factor or show irreducibility of polynomials.

determining a zero of a polynomial.
Let K be a field and $f \in K[X]$. Any polynomial of degree one in $K[X]$ is (up to a unit) of the form $X - a$ for some $a \in K$. By III.5.2 $X - a$ is a factor of f if and only if a is a zero of f. Searching for degree one factors of f is therefore equivalent to searching for zeros of f in K. The next three remarks can be useful here.

1. if $f = aX^2 + bX + c$ and $a \neq 0$ then

$$4a \cdot f = (2aX + b)^2 - (b^2 - 4ac)$$

('completing the square'). If $2 \neq 0$ in K one concludes that f has a zero in K if and only if $b^2 - 4ac$ is a square in K. The condition $2 \neq 0$ in K is necessary here since otherwise $4af = 0$. Note that for example in the field $K = \mathbb{F}_2$ we have $2 = 0$. The polynomial $X^2 + X + 1 \in \mathbb{F}_2[X]$ (so with $a = b = c = 1$) has $b^2 - 4ac = 1$ which is a square in $\mathbb{F}_2$; however the polynomial is irreducible.
2. in case the field K is *finite* one can try the elements of K one by one. As an example: $K = \mathbb{F}_3$, $f = X^3 + X + \overline{1}$; here $f(\overline{0}) = \overline{1}, f(\overline{1}) = \overline{0}, f(\overline{2}) = \overline{11} = \overline{2}$. Hence $\overline{1}$ is the only zero of f in K. One finds $f = (X - \overline{1})(X^2 + X - \overline{1})$.
3. in case $K = \mathbb{Q}$ (the field of fractions of the unique factorization domain $\mathbb{Z}$), we may assume that f is primitive in $\mathbb{Z}[X]$, so

$$f = a_n X^n + \ldots + a_1 X + a_0, \quad a_i \in \mathbb{Z},\ a_n \neq 0,\ \gcd(a_n, \ldots, a_1, a_0) = 1.$$

Moreover, factoring out the divisor(s) X if any, we may also assume $a_0 \neq 0$.
Claim: *every zero of f has the form $\frac{b}{c}$, with $b \in \mathbb{Z}$ a divisor of a_0 and $c \in \mathbb{Z}$ a divisor of a_n.*

Proof. Suppose $\frac{b}{c}$ is a zero of f, with $b, c \in \mathbb{Z}$ and $\gcd(b, c) = 1$. Then $f = (cX - b) \cdot g$ with $g \in \mathbb{Q}[X]$, and since $cX - b$ is primitive in $\mathbb{Z}[X]$ even $g \in \mathbb{Z}[X]$. Comparing the leading and the constant coefficients it follows that $c | a_n$ and $b | a_0$, as desired. ∎

Example: $f = 2X^3 + X^2 - X + 3$. The possibilities for b are $\pm 1, \pm 3$ and for c (which we may assume to be positive) only 1 and 2. Trying these eight possible fractions b/c one finds that $-3/2$ is the only zero of f in $\mathbb{Q}$.
An important special case is when $f \in \mathbb{Z}[X]$ is monic ($a_n = 1$). Then necessarily $c = 1$, so every rational zero is an *integer* and in fact a divisor of a_0.
In many cases one can reduce the number of possible fractions b/c even further by considering the *sign* of $f(x)$ or by considering the polynomial modulo a small prime number.

Example: $f = X^3 + X^2 + X + 6$. Possibilities for b/c: $\pm 1, \pm 2, \pm 3, \pm 6$. Clearly here $x > 0 \Rightarrow f(x) > 0$, and x odd $\Rightarrow f(x)$ odd. Hence only -2 and -6 need to be considered. Of these, only -2 turns out to be a zero of f.

Reducing modulo a prime number.
Claim: *Let $f \in \mathbb{Z}[X]$ be monic and suppose that a prime number p exists such that $(f \bmod p) \in \mathbb{F}_p[X]$ is irreducible. Then f is irreducible in $\mathbb{Q}[X]$ and also in $\mathbb{Z}[X]$.*

Proof. A factorization $f = g \cdot h$ in $\mathbb{Z}[X]$ yields $\overline{f} = \overline{g} \cdot \overline{h}$ (with $\overline{f} = (f \bmod p)$ in $\mathbb{F}_p[X]$), a contradiction. So f is irreducible in $\mathbb{Z}[X]$. By the lemma of Gauss (Corollary V.4.11) then also in $\mathbb{Q}[X]$. ∎

Example: $f = X^4 + 3X^3 - X^2 - X + 27$. Take $p = 2$, then $\overline{f} = X^4 + X^3 + X^2 + X + \overline{1}$ is irreducible in $\mathbb{F}_2[X]$ since it has no zero in $\mathbb{F}_2$ and it is not divisible by the only irreducible polynomial of degree 2 in $\mathbb{F}_2[X]$, which is $X^2 + X + \overline{1}$. Hence f is irreducible in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$.

Also in cases where $\overline{f}$ is *not* irreducible this method yields information.
Example: $f = X^4 - X^2 + X + 2$. Using a method explained earlier one verifies that f has no zero in $\mathbb{Q}$. Hence if f is reducible in $\mathbb{Z}[X]$ then $f = g \cdot h$ with g, h both of degree 2. So $\overline{f} = \overline{g} \cdot \overline{h}$ in $\mathbb{F}_2[X]$. However $\overline{f} \in \mathbb{F}_2[X]$ splits into the irreducible factors X and $X^3 + X + \overline{1}$, so $\overline{f}$ does not have factors of degree 2. Conclusion: f is irreducible in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$.

Eisenstein's criterion.
The irreducibility criterion we now discuss is named after the German mathematician Gotthold Eisenstein, 1823-1852.

V.5.1 Definition. Let R be a unique factorization domain and $p \in R$ irreducible.

$$f = a_n X^n + \ldots + a_1 X + a_0 \in R[X]$$

is called an *Eisenstein polynomial* (for p) if:

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

V.5.2 Theorem. *For a unique factorization domain R with field of fractions K, an Eisenstein polynomial $f \in R[X]$ for an irreducible $p \in R$ is irreducible in $K[X]$ and, in case f is primitive, also in $R[X]$.*

Proof. Since $\mathrm{co}(f)$ is not divisible by p, the primitive polynomial $f/\mathrm{co}(f)$ is an Eisenstein polynomial as well. Without lack of generality we will therefore assume that f is primitive. Suppose

$$f = g \cdot h, \qquad g, h \in R[X], \ \deg(g) > 0, \ \deg(h) > 0.$$

The assumptions imply that in $(R/pR)[X]$ we have

$$\overline{f} = (f \bmod p) = \overline{a}_n X^n \quad \text{and} \quad \overline{a}_n = (a_n \bmod p) \neq 0.$$

Moreover

$$\overline{f} = \overline{g} \cdot \overline{h}, \quad \deg(\overline{g}) > 0, \deg(\overline{h}) > 0.$$

This is only possible if

$$\overline{g} = \overline{b} X^k, \qquad \overline{h} = \overline{c} X^\ell$$

for some $b, c \in R$ and $k, \ell \in \mathbb{Z}_{>0}$. Hence the constant coefficients of g and h are both divisible by p, which implies that the constant coefficient a_0 of f is divisible by p^2, a contradiction. The primitive polynomial f is therefore irreducible in $R[X]$ hence by Corollary V.4.10 also in $K[X]$. This finishes the proof. ∎

V.5.3 Examples. $R = \mathbb{Z}, f = X^5 + 2X^3 - 6$; this is an Eisenstein polynomial for $p = 2$, hence it is irreducible in $\mathbb{Q}[X]$ and in $\mathbb{Z}[X]$.

$R = \mathbb{R}[Y], f = X^3 + (Y^4 - 1)X - (Y^2 + 1)$: this is an Eisenstein polynomial for $p = Y^2 + 1$. It is also primitive, hence irreducible in $\mathbb{R}[X, Y]$. The same holds for $X^2 + Y^2 - 1 \in (\mathbb{R}[Y])[X]$ using $p = Y - 1$.

Comparing coefficients.

If, as an example, we want to factor $\sum_{i=0}^{4} a_i X^i$ in $\mathbb{Z}[X]$ (with $a_0 \neq 0$, and we know that no factor of degree ≤ 1 exists, then we try

$$\sum_{i=0}^{4} a_i X^i = (b_2X^2 + b_1X + b_0)\cdot(c_2X^2 + c_1X + c_0).$$

Comparing coefficients implies in this case that

i. $b_2c_2 = a_4$
ii. $b_2c_1 + b_1c_2 = a_3$
iii. $b_2c_0 + b_1c_1 + b_0c_2 = a_2$
iv. $b_1c_0 + b_0c_1 = a_1$
v. $b_0c_0 = a_0$.

For b_2, c_2, b_0, c_0 using i. and v. there are only finitely many possibilities. For fixed b_2, c_2, b_0, c_0 we know b_1c_1 using iii., et cetera. This method is usually quite time consuming, but in the case of degree 4 it will lead in finitely many steps to a splitting of f in irreducible factors.

V.5.4 Remark. In the textbook *B.L. Van der Waerden, Algebra* (Volume I Chapter 5 §6 (Springer-Verlag, original German text 1931, this English translation 1991), an algorithm is presented which factors in finitely many steps any $f \in \mathbb{Z}[X]$ into irreducible elements. It is mainly of a theoretical value. For further (more practical) literature we refer to Chapter 2 of the book *H.G. Zimmer, Computational Problems, Methods, and Results in Algebraic Number Theory* (Springer-Verlag, 1972) or Chapter 3 §5 of *H. Cohen, A Course in Computational Number Theory* (Springer-Verlag, 1993)).

V.6 Exercises

1. We consider the ring

$$R = \mathbb{Z}[\sqrt{-5}] = \left\{a + b\sqrt{-5} : \ a, b \in \mathbb{Z}\right\}$$

   (a) Show that $2, 3 \in R$ are irreducible.
   Hint: use, as in I.2.5, the norm map

$$N : R \to \mathbb{Z}, \quad N(a + b\sqrt{-5}) = a^2 + 5b^2$$

   which has the property

$$r \in R^\times \Leftrightarrow N(r) = \pm 1.$$

   (b) Show that $R2$ and $R3$ are *not* prime ideals in R. Is R a unique factorization domain?

   (c) Why is the above not a contradiction with Theorem V.2.4 ?

   (d) Show that $6 = 2 \cdot 3$ and $6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ are two genuinely distinct factorizations of 6 as a product of irreducible elements of R.

2. Let R be the ring of polynomial functions on the circle:

$$R = \mathbb{R}[X, Y]/I, \qquad I = (X^2 + Y^2 - 1),$$

   and put $x := X + I, \ y := Y + I \in R$.

   (a) Prove that $x-1$ and $y-1$ are irreducible in R. (Hint: use the map $N : R \to \mathbb{R}[X]$ given in Exercise 18 on page 51).

   (b) Prove that $(x-1)$ and $(y-1)$ are not prime ideals in R and that R is not a unique factorization domain.

   (c) Show that $a = (x + y - 1)^2 = 2(x-1)(y-1)$ are two distinct factorizations of a as a product of irreducible elements (and a unit 2).

   (d) Draw a picture of the circle and the lines $X + Y - 1 = 0$, $X - 1 = 0$, and $Y - 1 = 0$. Find some more elements in R which allow two distinct factorizations into irreducible elements.

3. For each of the following elements of $\mathbb{Z}[\sqrt{-3}]$, determine whether it is irreducible and whether it generates a prime ideal:

$$\sqrt{-3}, \ 1, \ 2, \ 1 + \sqrt{-3}, \ 5.$$

4. Let $R = \{\sum a_i X^i \in \mathbb{Q}[X] : a_1 = 0\}$ as in Example V.1.5.

   (a) Let

$$\mathrm{ev}_0 : R \longrightarrow \mathbb{Q}, \qquad f \mapsto f(0)$$

   be the evaluation homomorphism in 0. Prove that

$$\ker(\mathrm{ev}_0) = (X^2, X^3) = \{f = X^2 g + X^3 h \in R : \ g, h \in R\}.$$

   (b) Prove that $\ker(\mathrm{ev}_0)$ is not a principal ideal, and that it is a maximal ideal.

5. Let $R = \{a/b \in \mathbb{Q} : \ a, b \in \mathbb{Z}, \ b \ \textit{odd}\ \}$. This is a subring of $\mathbb{Q}$.

   (a) Determine $R^\times$.

   (b) Prove that every $x \in R$, $x \neq 0$ can be written in a unique way as $x = 2^k \cdot u$ for some $k \in \mathbb{Z}_{\geq 0}$, $u \in R^\times$.

(c) Show that 2 is up to multiplication by units the unique irreducible element of R. Is $2R$ a prime ideal?

6. Let $R = \mathbb{Z}[X]/(5X, X^2)$.

   (a) Prove that every element of R can be written in a unique way as

   $$\overline{a} + \overline{b} \cdot \overline{X} \quad \text{with} \quad a \in \mathbb{Z},\ b \in \mathbb{Z},\ 0 \le b < 5.$$

   Here $^-$ denotes the residue class modulo $(5X, X^2)$.

   (b) Prove: $\overline{a} + \overline{bX} \in R^\times \iff a \in \{\pm 1\}$.

   (c) Prove: if $\alpha = \overline{X}$, $\beta = \overline{2} \cdot \overline{X}$ then

   $$R \cdot \alpha = R \cdot \beta \qquad \text{and} \quad \alpha \notin R^\times \cdot \beta.$$

7. Let R be a unique factorization domain and $a, b \in R$ not both zero. Put $d = \gcd(a, b)$. Suppose that $c \in R$ is a divisor of both a and b, so $a_1,\ b_1 \in R$ exist with $a = ca_1,\ b = cb_1$. Prove that c is a divisor of d.
8. Factor $X^8 - 16$ and $X^6 + 27$ into irreducible elements of $\mathbb{Q}[X]$.
9. Is $5X^4 + 10X + 10$ an Eisenstein polynomial in $\mathbb{Z}[X]$? Is it irreducible in $\mathbb{Z}[X]$? And in $\mathbb{Q}[X]$?
10. Show that $X^n + 2$ is irreducible in $\mathbb{Z}[X]$ for all $n \in \mathbb{Z}_{\ge 0}$.
    Prove that $Y^n - X$ is irreducible in $K[X, Y]$ (K a field) for all $n \in \mathbb{Z}_{\ge 0}$.
11. (a) Find an example of an irreducible polynomial $f \in \mathbb{Z}[X]$ with the property that $f(X^2)$ is *not* irreducible.

    (b) Let $f \in \mathbb{Z}[X]$ be a monic Eisenstein polynomial. Show that $f(X^2)$ is irreducible in $\mathbb{Z}[X]$.
12. Let R be a unique factorization domain. prove that

    $$\cup_{n \ge 0} R[X_1, X_2, \ldots, X_n]$$

    is a unique factorization domain as well.
13. Write the following polynomials as a product of irreducible elements in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$:

    $$\begin{aligned} 4X^2 + 4, \\ 2X^{10} + 4X^5 + 3, \\ X^4 - 7X^2 + 5X - 3, \\ X^{111} + 9X^{74} + 27X^{37} + 27, \\ X^3 + X + 3. \end{aligned}$$

14. Write the following polynomials as a product of irreducible elements in $\mathbb{Z}[X]$ and in $\mathbb{Q}[X]$:

    $$\begin{aligned} \tfrac{1}{7}((X+1)^7 - X^7 - 1), \\ X^3 + 3X^2 + 6X + 9, \\ X^4 + 2X^3 + 3X^2 + 9X + 6, \\ X^{12} - 1, \\ X^4 - X^3 + X^2 - X + 1. \end{aligned}$$

15. Write the following polynomials as a product of irreducible elements in $\mathbb{Q}[X, Y]$:

    $$\begin{aligned} Y^4 + X^2 + 1, \\ Y^3 - (X+1)Y^2 + Y + X(X-1), \\ X^n + Y^3 + Y \quad (n \ge 1), \\ X^4 + 4Y^4, \\ X^4 + 2X^3 + X^2 - Y^2 - 2Y - 1, \\ Y^n - 13X^4 \qquad (n \ge 1). \end{aligned}$$

16. Let $I \subset \mathbb{Z}[X]$ be a prime ideal.
    (a) Prove that $I \cap \mathbb{Z}$ is a prime ideal in $\mathbb{Z}$.
    (b) Prove that either $I = \{0\}$ or $I = (f)$ for some irreducible $f \in \mathbb{Z}[X]$ or $I = (p)$ with $p \in \mathbb{Z}$ a prime number or $I = (p, f)$ where $f \in \mathbb{Z}[X]$ is irreducible modulo the prime number p.
    (c) Describe the maximal ideals of $\mathbb{Z}[X]$.
17. Suppose that $n > 0$ is an integer such that $n^4 + 4^n$ is a prime number. Prove that $n = 1$.
18. Let $f \in \mathbb{Z}[X]$ be a monic polynomial such that $f(0)$ is a prime number. Prove that f has at most *three* distinct zeros in $\mathbb{Q}$.
19. Find all irreducible $f \in \mathbb{F}_2[X]$ of degree at most 3.
20. Let $R = \mathbb{C}[U, V]/(UV - 1)$.
    (a) Prove that
    $$\mathbb{C}[T, T^{-1}] := \left\{ \frac{f(T)}{T^i} \in \mathbb{C}(T) : \ f(T) \in \mathbb{C}[T],\, i \in \mathbb{Z} \right\}$$
    is a subring of the field $\mathbb{C}(T)$.
    (b) Prove that $R \cong \mathbb{C}[T, T^{-1}]$.
    (c) Prove that R is a principal ideal domain.
    (d) Prove that $R \cong \mathbb{C}[X, Y]/(X^2 + Y^2 - 1)$ (hint: $X^2 + Y^2 = (X + iY)(X - iY)$).
    (e) Find an element $r \in \mathbb{C}[X, Y]/(X^2 + Y^2 - 1)$ such that $(r) = (x - 1, y)$ where $x := X + (X^2 + Y^2 - 1)$ and $y = Y + (X^2 + Y^2 - 1)$. (compare Exercise 18 on page 51).

VI EUCLIDEAN RINGS AND THE GAUSSIAN INTEGERS

VI.1 Euclidean rings

We now discuss a special kind of principal ideal domains. Examining how we determined that the rings $\mathbb{Z}$ and $K[X]$ (K a field) are indeed principal ideal domains, we see that in both cases an important rôle was played by the possibility to perform *division with remainder*: compare Theorem II.4.2 for the ring $\mathbb{Z}$, and Theorem III.4.1 for $K[X]$. Integral domains allowing such a division with remainder are called *Euclidean*. We start by formally introducing them.

VI.1.1 Definition. A ring R is called a *Euclidean ring* if it is a domain, and moreover a function

$$g : R - \{0\} \longrightarrow \mathbb{Z}_{\geq 0}$$

exists with the property:
(**) for all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ with $a = qb + r$ and either $r = 0$ or $g(r) < g(b)$.

VI.1.2 Remark. Property (**) expresses the possibility to do 'division with remainder'. The function g is used to say that the 'remainder' r is *smaller* than the element b one divides by.

VI.1.3 Example. For $R = \mathbb{Z}$ one can take $g(n) = |n|$, in which case the familiar division with remainder for integers (and the fact that $\mathbb{Z}$ is an integral domain) shows that $\mathbb{Z}$ is Euclidean.

Next, take K any field and $R = K[X]$. Then $g(f) = \deg(f)$ works, as follows from Theorem III.3.1. Moreover $K[X]$ is an integral domain, hence $K[X]$ is Euclidean.

A field K is trivially Euclidean: we may take $g(a) = 0$ for all $a \in K - \{0\}$. ∎

VI.1.4 Theorem. *Any Euclidean ring R is a principal ideal domain.*

Proof. By definition R is a domain. Take $I \subset R$ any ideal. We will show that I is a principal ideal. If $I = \{0\}$ this is clear. Now suppose $I \neq \{0\}$. Then $I - \{0\}$ is not empty, hence $g(I - \{0\})$ is a nonempty subset of $\mathbb{Z}_{\geq 0}$. This subset contains a smallest element, hence $b \in I - \{0\}$ exists with

$$g(b) = \min\{g(x) : \ x \in I - \{0\}\}.$$

We claim that

$$I = Rb.$$

The inclusion $\supset$ is clear since $b \in I$. The inclusion $\subset$ is shown as follows. Let $a \in I$. Since R is Euclidean, $q, r \in R$ exist with $a = qb + r$ and either $r = 0$ or $g(r) < g(b)$. If $r \neq 0$ then $g(r) < g(b)$, and $r = a - qb \in I$. This contradicts the choice of b. As a consequence, $r = 0$, which implies $a = qb \in Rb$, as we wished to show.

We conclude that $I = Rb$ is a principal ideal, which proves Theorem VI.1.4. ■

Note that the above proof is completely analogous to the proofs of II.4.2 (for $R = \mathbb{Z}$) and III.4.1 (for $R = K[X]$).

The converse of VI.1.4 is not true in general: there exist principal ideal domains which are not Euclidean. An example of this is the ring $R = \mathbb{Z}[\frac{1}{2}(1 + \sqrt{-19})]$, see Exercises 2 and 3 on page 89.

The types of rings discussed so far are displayed as follows.

$$\{\text{Rings}\} \supsetneq \{\text{Comm. Rings}\} \supsetneq \{\text{ID's}\} \supsetneq \{\text{UFD's}\} \supsetneq \{\text{PID's}\} \supsetneq \{\text{EuD's}\} \supsetneq \{\text{Fields}\}.$$

Here the following (mostly quite standard) abbreviations were used

Comm. Rings	commutative rings	(Definition I.1.1)
ID's	integral domains	(Definition I.2.13)
UFD's	unique factorization domains	(Definition V.3.1)
PID's	principal ideal domains	(Definition V.2.1)
EuD's	Euclidean rings	(Definition VI.1.1)

VI.1.5 Theorem. *The ring $\mathbb{Z}[i]$ is a Euclidean ring, taking g the norm map*

$$g(a+bi) := N(a+bi) = (a+bi)\overline{(a+bi)} = a^2 + b^2, \quad \text{for } a, b \in \mathbb{Z}.$$

In particular $\mathbb{Z}[i]$ is a principal ideal domain.

Proof. Take $\alpha, \beta \in \mathbb{Z}[i], \beta \neq 0$. We must find $\gamma, \rho \in \mathbb{Z}[i]$ such that

$$\alpha = \gamma\beta + \rho \qquad \text{and} \quad N(\rho) < N(\beta)$$

(note that $N(0) = 0$). So in $\mathbb{C}$ we must have $\alpha/\beta = \gamma + \rho/\beta$ and $N(\rho/\beta) < 1$ (for the obvious extension of the norm N to all of $\mathbb{C}$). This may be interpreted as stating that $\gamma \in \mathbb{Z}[i]$ should be a good approximation of the complex number α/β.

In $\mathbb{C}$ write

$$\frac{\alpha}{\beta} = u + vi, \qquad \text{with} \quad u, v \in \mathbb{R}.$$

Choose

$$u', v' \in \mathbb{Z} \quad \text{such that} \quad |u - u'| \leq \frac{1}{2} \quad \text{and} \quad |v - v'| \leq \frac{1}{2}.$$

Put

$$\gamma = u' + v'i \in Z[i]$$

and define the 'remainder' ρ by

$$\rho := \alpha - \gamma\beta \in \mathbb{Z}[i], \qquad \text{so} \qquad \alpha = \gamma\beta + \rho.$$

We claim that in this way a (not necessarily unique) division with remainder as required in the definition of a Euclidean ring is obtained.

Indeed, since $N(\alpha)N(\beta) = N(\alpha\beta)$, the inequality $N(\rho) < N(\beta)$ follows from

$$\begin{aligned} N(\rho/\beta) &= N(\frac{\alpha}{\beta} - \gamma) \\ &= N((u-u') + (v-v')i) \\ &= (u-u')^2 + (v-v')^2 \\ &\leq (\frac{1}{2})^2 + (\frac{1}{2})^2 \\ &= \frac{1}{2} \\ &< 1. \end{aligned}$$

This shows that $\mathbb{Z}[i]$ is a Euclidean ring. Theorem VI.1.4 now implies that $\mathbb{Z}[i]$ is a principal ideal domain. ∎

VI.1.6 Example. We do division with remainder for $\alpha = 5+6i$ and $\beta = 2+i$. First, in $\mathbb{C}$ one has

$$\frac{\alpha}{\beta} = \frac{(5+6i)(2-i)}{(2+i)(2-i)} = \frac{16+7i}{5} = 3\frac{1}{5} + i \cdot 1\frac{2}{5}.$$

Here a 'good approximation' γ for α/β is

$$\gamma := 3+i \quad \Longrightarrow \rho := \alpha - \gamma\beta = 5+6i-(3+i)(2+i) = i.$$

Hence a division with remainder as requested is

$$5+6i = (3+i)(2+i) + i, \qquad \text{note that indeed} \quad N(i) = 1 < 5 = N(2+i).$$

∎

The proof of Theorem VI.1.5 presented above can be interpreted geometrically as follows: we must show that any complex number α/β can be approximated by an element of $\mathbb{Z}[i]$ in such a way that the difference has absolute value < 1. In other words: the open discs with radius 1 and center the elements of $\mathbb{Z}[i]$, should cover the complex plane. The fact that they indeed do is immediate after drawing a picture.

There are various rings where an analogous reasoning shows that they are Euclidean. Examples include $\mathbb{Z}[\sqrt{-2}]$ and $\mathbb{Z}[\frac{1}{2}(1+\sqrt{-3})] = \{a + \frac{1}{2}(1+\sqrt{-3})b : a, b \in \mathbb{Z}\}$. In the complex plane, this latter ring consists of the vertices of a regular pattern consisting of equilateral triangles. The fact that this ring is Euclidean can be used to prove *'Fermat's Last Theorem* in the special case $n = 3$. The general case of this theorem states that no $x, y, z \in \mathbb{Z}_{>0}$ exist with $x^n + y^n = z^n$, in case n is an integer larger than 2. Fermat claimed to have a proof of this, but it is unknown (and strongly doubted) whether that was indeed the case. For centuries many mathematicians and amateurs attempted to prove Fermat's Last Theorem. In June 1993 the English mathematician Andrew Wiles (Who worked at Princeton University in the U.S.A.) announced that he, building upon the work of a long list of number theorists and algebraic geometers, finally succeeded in this. Unfortunately this first 'proof' still contained a serious flaw, but little more than a year later Wiles in collaboration with his former student Richard Taylor succeeded in completing the argument. The now fully accepted proof appeared in the journal Annals of Mathematics **142** (1995).

The ring $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain as can be seen using Exercise 17 on page 32. hence this ring is certainly not Euclidean. So the discs of radius 1 centered around the elements of $\mathbb{Z}[\sqrt{-5}]$ do not cover the complex plane. In a similar way one shows that $Z[\sqrt{-3}]$ is not a principal ideal domain, hence not a

Euclidean ring. In this example it turns out that the region of the complex plane which is not covered by the discs, consists of a set of isolated points.

Also for $m > 0$ there are Euclidean rings of the form $\mathbb{Z}[\sqrt{m}]$. Examples include the rings $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[\sqrt{3}]$, with

$$(*) \qquad g(a) = |N(a)|, \qquad N \text{ as in } I.2.5.$$

For more examples we refer to the Exercises. The proofs are in these cases quite analogous to the proof presented for $\mathbb{Z}[i]$.

An interesting ring is $\mathbb{Z}[\sqrt{14}]$: here the function defined by $(*)$ does not satisfy the condition needed to show the ring is Euclidean. However, it is believed that some other function g exists which *does* fulfill the requirement. However, to date this remains unproven, so we do not know whether $\mathbb{Z}[\sqrt{14}]$ is Euclidean. We do know that $\mathbb{Z}[\sqrt{14}]$ is a principal ideal domain.

VI.2 The Euclidean algorithm

Throughout this section R is a Euclidean ring.

In a principal ideal domain (such as R, see Theorem VI.1.4) for every pair of elements a, b the ideal (a, b) generated by them is principal. So d in the ring exists such that

$$(a, b) = (d), \qquad \text{and in particular} \qquad ar + bs = d$$

for certain r, s in the ring.

VI.2.1 Definition. Given elements a, b in a principal ideal domain, a generator d of the principal ideal (a, b) is called a *greatest common divisor* of a and b. We write $\gcd(a, b) := d$.

VI.2.2 Remark. Note that in general a greatest common divisor d of a, b is not uniquely determined. If u is any unit in the ring, then $(d) = (ud)$; hence ud is also a greatest common divisor of a and b.

VI.2.3 Remark. From Theorem V.3.4 we have that a principal ideal domain is in particular a unique factorization domain. In unique factorization domains we already defined a gcd in Definition V.4.5. Exercise 11 on page 90 shows that the two definitions agree.

In a Euclidean ring there is a method called the *Euclidean algorithm* for finding a greatest common divisor of any pair of elements.

Given a, b in a Euclidean ring R. We assume both a and b are nonzero, otherwise we already have a gcd. Assume (after interchanging a, b if necessary) that $g(b) \leq g(a)$. Division with remainder yields $q_0, r_1 \in R$ such that

$$a = q_0 b + r_1 \qquad \text{and} \quad r_1 = 0 \quad \text{or} \quad g(r_1) < g(b).$$

Since $a, b \in (a, b)$ also $r_1 = a - q_0 b \in (a, b)$, and

$$(a, b) = (q_0 b + r_1, b) = (b, r_1).$$

If $r_1 = 0$ then $(a, b) = (q_0 b, b) = (b)$ hence $\gcd(a, b) = b$. And if $r_1 \neq 0$ we have $g(r_1) < g(b) \leq g(a)$, hence we obtained 'smaller' generators b, r_1 of the ideal (a, b).

If $r_1 \neq 0$ one continues by dividing b with remainder by r_1:

$$b = q_1 r_1 + r_2, \qquad \text{and} \quad r_2 = 0 \text{ or } g(r_2) < g(r_1).$$

Then

$$(a, b) = (b, r_1) = (q_1 r_1 + r_2, r_1) = (r_1, r_2).$$

If $r_2 \neq 0$ then divide r_2 by r_1:

$$r_1 = q_2 r_2 + r_3 \qquad \text{and} \quad r_3 = 0 \text{ or } g(r_3) < g(r_2),$$

it follows that $(a, b) = (b, r_1) = (r_1, r_2) = (r_2, r_3)$.

Note that in each new step either $r_k = 0$ or $g(r_k) < g(r_{k-1})$, and $g(r_k) \in \mathbb{Z}_{\geq 0}$. So after finitely many steps one obtains n with

$$r_{n-2} = q_{n-1} r_{n-1} + r_n \qquad \text{and} \quad r_n = 0.$$

Then

$$(a, b) = (r_{n-1}, r_n) = (r_{n-1}), \qquad \text{hence} \quad \gcd(a, b) = r_{n-1},$$

which determines a gcd of a and b.

Elements $r, s \in R$ with $ar + bs = d$ are now constructed as follows.

$$\left.\begin{array}{rcl} a - q_0 b & = & r_1 \\ b - q_1 r_1 & = & r_2 \end{array}\right\} \Longrightarrow b - q_1(a - q_0 b) = r_2, \quad \text{i.e.,} \quad (-q_1)a + (1 + q_0 q_1)b = r_2,$$

where we substituted the first equality into the second. Now if

$$\left.\begin{array}{rcl} h_{i-1}a + \quad k_{i-1}b & = & r_{i-1} \\ h_i a + \quad k_i b & = & r_i \end{array}\right\} \qquad \text{and} \quad r_{i-1} - q_i r_i = r_{i+1},$$

then a substitution shows

$$(h_{i-1} - q_i h_i)a + (k_{i-1} - q_i k_i)b = r_{i+1},$$

so

$$h_{i+1} = (h_{i-1} - q_i h_i), \quad k_{i+1} = (k_{i-1} - q_i k_i)$$

satisfy $h_{i+1}a + k_{i+1}b = r_{i+1}$. After finitely many steps in this way the desired expression for a gcd is obtained.

A different way to keep track of this 'bookkeeping' is as follows. Put

$$r_{-1} := a, \qquad r_0 := b.$$

Now observe that the equation $r_{i-1} = q_i r_i + r_{i+1}$ is equivalent with

$$\begin{pmatrix} r_i \\ r_{i+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_i \end{pmatrix} \begin{pmatrix} r_{i-1} \\ r_i \end{pmatrix}.$$

Considering the second coordinate of the vector

$$\begin{pmatrix} r_{n-2} \\ r_{n-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_{n-2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -q_{n-3} \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} 0 & 1 \\ 1 & -q_0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

one finds r, s such that $\gcd(a, b) = r_{n-1} = ra + bs$.

VI.2.4 Example. We use the algorithm described above to find $\gcd(84, 30)$ in the euclidean ring $\mathbb{Z}$ (so $g(n) = |n|$), and write the result as a combination $84r + 30s$. Note

$$84 = 2 \cdot 30 + 24, \qquad 30 = 1 \cdot 24 + 6, \qquad 24 = 4 \cdot 6 + 0,$$

so $(84, 30) = (30, 24) = (24, 6) = (6)$ and $\gcd(84, 30) = 6$. Moreover

$$24 = 84 - 2 \cdot 30, \quad 30 - 1 \cdot 24 = 6 \quad \Longrightarrow \quad 30 - 1 \cdot (84 - 2 \cdot 30) = (-1) \cdot 84 + 3 \cdot 30 = 6,$$

hence $r = -1$ and $s = 3$ work. ∎

VI.2.5 Example. The Euclidean algorithm can be used to compute the inverse of a unit in the ring $R/(a)$. We illustrate this in an example, see also VII.2.10.

Let $R = \mathbb{Q}[X]$ (a Euclidean ring), and

$$f = X^3 + X^2 - 1 \quad \in \mathbb{Q}[X].$$

Then f is irreducible since it has no zero in $\mathbb{Z}$ and therefore not in $\mathbb{Q}$ (compare Section V.5). Put

$$K := \mathbb{Q}[X]/(f), \qquad \alpha := X + (f).$$

Using Theorem V.2.4 and Theorem IV.2.3 K is a field. Moreover by Theorem III.3.4 every element of K can be written in a unique way as

$$a_0 + a_1\alpha + a_2\alpha^2, \qquad a_i \in \mathbb{Q}.$$

We determine the inverse of

$$b(\alpha) = 1 + \alpha^2, \qquad \text{for} \quad b = X^2 + 1 \quad (\in \mathbb{Q}[X]).$$

Since $\mathbb{Q}[X]$ is a principal ideal domain and f is irreducible with $\deg(f) > \deg(b)$, we have $\gcd(f, b) = 1$. Hence $r, s \in \mathbb{Q}[X]$ exist with

$$fr + sb = 1 \quad (\in \mathbb{Q}[X]), \qquad \text{so} \quad s(\alpha)(\alpha^2 + 1) = 1 \quad (\in K = \mathbb{Q}[X]/(f)),$$

using $f(\alpha) = 0$. So $s(\alpha)$ is the inverse of $b(\alpha) = \alpha^2 + 1$.

As $\mathbb{Q}[X]$ is a Euclidean ring (with $g(\cdot) = \deg(\cdot)$), we can find s using the Euclidean algorithm, as follows. We have

$$X^3 + X^2 - 1 = (X + 1)(X^2 + 1) + (-X - 2), \qquad \text{so} \quad q_0 = X + 1, \quad r_1 = -(X + 2).$$

Furthermore

$$X^2 + 1 = (-X + 2)(-X - 2) + 5, \qquad \text{hence} \quad q_1 = -X + 2, \quad r_2 = 5.$$

Since 5 is a unit in $\mathbb{Q}[X]$, indeed $\gcd(f, b) = 1$.

These equalities can be rewritten as

$$-X - 2 = f - (X + 1)b, \qquad 5 = b + (X - 2)(-X - 2).$$

Substituting the first into the second yields

$$5 = b + (X - 2)(f - (X + 1)b) = (X - 2)f + (1 - (X - 2)(X + 1))b.$$

So one obtains

$$rf + sb = 1 \qquad \text{with} \qquad r = \frac{1}{5}(X - 2), \quad s = \frac{1}{5}(3 + X - X^2).$$

Substituting α for X we see that

$$(\alpha^2 + 1)^{-1} = \frac{1}{5}(3 + \alpha - \alpha^2).$$

VI.3 Sums of squares

In this section we answer the question, which integers can be written as a sum of two squares. For this we use the theory of unique factorization domains and

a result concerning finite subgroups of the group of units of a domain, see Corollary III.5.4. Moreover for actually *computing* a representation as a sum of two squares, the Euclidean algorithm can play a role.

Writing $n \in \mathbb{Z}$ as a sum of two squares, $n = a^2 + b^2$ with $a, b \in \mathbb{Z}$, can be considered as factoring n in the ring $\mathbb{Z}[i]$:

$$n = a^2 + b^2 \quad \Longleftrightarrow \quad n = (a + bi)(a - bi).$$

In particular any factorization of n in $\mathbb{Z}[i]$ into two complex conjugate factors $n = \alpha \cdot \overline{\alpha}$ yields a a representation of n as a sum of two squares (one of the squares could be 0).

VI.3.1 Definition. The ring $\mathbb{Z}[i]$ is called the ring of *Gaussian integers*.

We saw in Theorem VI.1.5 that $\mathbb{Z}[i]$ is Euclidean, hence in particular it is a principal ideal domain and therefore also a unique factorization domain. As a consequence, every $n \in \mathbb{Z}_{>0}$ has in $\mathbb{Z}[i]$ an (essentially unique) factorization as a product of a unit and a finite number of irreducible elements of $\mathbb{Z}[i]$.

VI.3.2 Example. Observe that 5 is irreducible in $\mathbb{Z}$, however in $\mathbb{Z}[i]$

$$5 = 1^2 + 2^2 = (1 + 2i)(1 - 2i)$$

and neither of $1 + 2i$, $1 - 2i$ is a unit. So 5 is *not* irreducible in $\mathbb{Z}[i]$. ∎

All α's in $\mathbb{Z}[i]$ such that $\alpha\overline{\alpha} = n$ can be found once we know the factorization of n into irreducible elements of $\mathbb{Z}[i]$: since α divides n, its irreducible factors form a subset of those of n.

Let

$$n = p_1^{n_1} \dots p_t^{n_t}, \qquad n_j \in \mathbb{Z}_{>0}$$

and the p_j pairwise distinct prime numbers. If we know how to write each p_j as a product of irreducible elements in $\mathbb{Z}[i]$, then the factorization of n into irreducible elements of $\mathbb{Z}[i]$ is obtained.

VI.3.3 Theorem. *1. The units of* $\mathbb{Z}[i]$ *are* $\mathbb{Z}[i]^\times = \{1, i, -1, -i\}$.
2. *One has* $2 = (-i)(1 + i)^2$ *with* $-i \in \mathbb{Z}[i]^\times$ *and* $(1 + i)$ *irreducible in* $\mathbb{Z}[i]$.
3. *If* $q \in \mathbb{Z}$ *is prime and* $q \equiv 3 \bmod 4$ *then* q *is irreducible in* $\mathbb{Z}[i]$.
4. *If* $p \in \mathbb{Z}$ *is prime and* $p \equiv 1 \bmod 4$ *then*

$$p = \pi \cdot \overline{\pi}, \qquad \text{and} \quad \pi \neq u\bar{\pi}$$

for any $u \in \mathbb{Z}[i]^\times$. *Both* π *and its complex conjugate* $\overline{\pi}$ *are irreducible in* $\mathbb{Z}[i]$.
5. *The irreducible elements mentioned in 2. and 3. and 4. above are, up to multiplication by units, the only irreducible elements of* $\mathbb{Z}[i]$.

Proof. 1.) In I.2.5 we saw $a + bi \in \mathbb{Z}[i]^\times \iff N(a + bi) := a^2 + b^2 = \pm 1$, implying 1).
2.) Note that

$$N(1 + i) = 1^2 + 1^2 = 2 \quad \text{and} \quad N(\alpha)N(\beta) = N(\alpha\beta).$$

Hence if $\alpha\beta = 1 + i$ then either $N(\alpha) = 1$ or $N(\beta) = 1$. As shown above, this means that one of α, β is a unit. Hence $1 + i$ is irreducible in $\mathbb{Z}[i]$.
3.) Suppose $q = \alpha\beta$ with neither α nor β a unit. Then

$$N(\alpha)N(\beta) = q^2, \qquad N(\alpha) > 1, \quad N(\beta) > 1.$$

Since $q \in \mathbb{Z}$ is prime, this implies $N(\alpha) = N(\beta) = q$. Write

$$\alpha = a + bi, \qquad \text{then} \quad N(\alpha) = a^2 + b^2 = q.$$

If a and b are both even then $a^2 + b^2 \equiv 0 \bmod 4$ contradicting $q \equiv 3 \bmod 4$. If a and b are both odd, then writing

$$a = 2k + 1 \qquad \text{one finds} \quad a^2 = 4k^2 + 4k + 1 \equiv 1 \bmod 4.$$

Hence in this case $a^2 + b^2 \equiv 2 \bmod 4$, contradicting the assumption on q. In the remaining case one of a, b is even and the other one is odd. Then $a^2 + b^2 \equiv 1 \bmod 4$, again a contradiction. Hence no $\alpha \in \mathbb{Z}[i]$ exists with $N(\alpha) = q$ and therefore q is irreducible.

4.) If p is any prime number, the ring $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a domain (even a field), hence by Corollary III.5.4 $\mathbb{F}_p^\times$ is a cyclic group of order $p - 1$. Let g be a generator of this group. If our prime $p \equiv 1 \bmod 4$, then $(p-1)/4$ is a positive integer, hence

$$x := g^{(p-1)/4} \qquad (\in \mathbb{F}_p^\times)$$

is well-defined. The order of x equals 4, so $x^2 = -1$. As a consequence the map

$$\phi : \mathbb{Z}[i] \longrightarrow \mathbb{F}_p, \qquad a + bi \mapsto \overline{a} + \overline{b}x$$

is a (surjective) ring homomorphism. By Theorem VI.1.5 $\mathbb{Z}[i]$ is a principal ideal domain, hence $\pi \in \mathbb{Z}[i]$ exists with

$$(\pi) := \pi\mathbb{Z}[i] = \ker(\phi), \qquad \text{in particular} \quad \mathbb{Z}[i]/(\pi) \cong \mathbb{F}_p.$$

Clearly $p \in \ker(\phi)$, so $p = \pi\beta$ for some $\beta \in \mathbb{Z}[i]$. Then $N(p) = p^2 = N(\pi)N(\beta)$. Here $N(\pi) \neq 1$ since otherwise π would be a unit and $(\pi) = \mathbb{Z}[i]$, contradicting $\mathbb{Z}[i]/(\pi) = \mathbb{F}_p$. If $N(\pi) = p^2$ then $N(\beta) = 1$ hence β is a unit and $(\pi) = (\pi\beta) = (p)$. This is impossible since $\mathbb{Z}[i]/(p)$ is a ring consisting of p^2 elements (the classes of $a + bi$ for $a, b \in \{0, 1, \ldots, p-1\}$) whereas $\mathbb{Z}[i]/(\pi) \cong \mathbb{F}_p$ contains only p elements. We conclude $N(\pi) = p$. Observe that $p = N(\pi) = \pi\overline{\pi}$ (hence $\beta = \overline{\pi}$), which yields a factorization of p.

The irreducibility of π (and of $\overline{\pi}$) follows from the remark that $\pi = \alpha\gamma$ implies that $N(\pi) = p = N(\alpha)N(\gamma)$, so $N(\alpha) = 1$ or $N(\gamma) = 1$ showing that one of α, γ is a unit.

We finish by showing that no unit u exists with $\pi = u\overline{\pi}$. Write $\pi = a + bi$ and recall $N(\pi) = p = a^2 + b^2$. Suppose

$$a + bi = u(a - bi) \qquad \text{with} \quad u \in \mathbb{Z}[i]^* = \{1, i, -1, -i\}.$$

If $u = \pm 1$ this yields $a = 0$ or $b = 0$, contradicting $p = a^2 + b^2$. If $u = \pm i$ then $a = \pm b$, contradicting the assumption that $p = a^2 + b^2 \equiv 1 \bmod 4$.

5.) If $\alpha \in \mathbb{Z}[i]$ is irreducible then $n := \alpha \cdot \overline{\alpha} \in \mathbb{Z}_{>1}$. Factoring n in $\mathbb{Z}$ we see that α is a divisor in $\mathbb{Z}[i]$ of some prime factor $p|n$. The irreducible factors of this p are as described in 2., 3., or 4.

This proves Theorem VI.3.3. ∎

VI.3.4 Remark. The hardest part of the reasoning above is probably the factorization of a prime $p \equiv 1 \bmod 4$ in $\mathbb{Z}[i]$. This can also be shown by more elementary means. In particular the use of Corollary III.5.4 can be avoided. For a quite short and elementary argument, see D. Zagier: *A one-sentence proof that every prime* $\equiv 1 \bmod 4$ *is a sum of two squares*, The American Mathematical Monthly, Vol. 97 (1990), p. 144.

VI.3.5 Remark. Here is a somewhat more constructive proof for the existence of a representation $p = a^2 + b^2$ whenever $p \equiv 1 \bmod 4$ is prime. First, consider

$$(p-1)! \bmod p = (1 \cdot 2 \cdot 3 \cdot \ldots \cdot (p-1)) \bmod p.$$

This is the product of all elements of $\mathbb{F}_p^\times$. With any factor $\overline{x}$ in this product, also the inverse $\overline{x}^{-1}$ is a factor. One has

$$\overline{x} = \overline{x}^{-1} \Leftrightarrow \overline{x}^2 = \overline{1} \Leftrightarrow \overline{x} = \pm\overline{1}.$$

Conclusion:

$$(p-1)! \equiv -1 \bmod p,$$

as is seen by grouping together all pairs $\overline{x}, \overline{x}^{-1}$ with $\overline{x} \neq \pm\overline{1}$. This result is named after the English mathematician John Wilson (1741–1793), although it was also stated much earlier by the Arab scientist Alhazen (Ibn al-Haytham, ≈965–≈1040 A.D.). If one subtracts p from each of the last $(p-1)/2$ factors $(p+1)/2, \ldots, (p-1)$, observing that $(p-j) - p = -j$ and that $(p-1)/2$ is *even* in our case, one obtains

$$(\tfrac{p-1}{2})!^2 \equiv -1 \bmod p,$$

hence $x := (\frac{p-1}{2})! \bmod p \in \mathbb{F}_p^\times$ satisfies $x^2 = -\overline{1}$.

Given $x \in \mathbb{F}_p$ with $x^2 = -\overline{1}$, take $n \in \mathbb{Z}$ with $|n| < p/2$ and $n \bmod p = \pm x$. The map $\phi : \mathbb{Z}[i] \to \mathbb{F}_p$ used in the proof of Theorem VI.3.3 may be written as

$$a + bi \mapsto \phi(a+bi) = \overline{a + bn}.$$

The kernel of this map consists, by definition, of all $a + bi \in \mathbb{Z}[i]$ with $p \mid (a + bn)$. In other words,

$$\ker(\phi) = \{a + bi \,:\, \exists m \in \mathbb{Z} \text{ such that } a + bn = mp\}.$$

Writing $a = mp - bn$ this shows that

$$\ker(\phi) = \{mp - bn + bi \,:\, m, b \in \mathbb{Z}\}$$

hence this kernel is generated by the two elements p and $n - i$. Since it is an *ideal* in $\mathbb{Z}[i]$, one concludes $\ker(\phi) = (p, n-i)$.

Now the Euclidean algorithm (Section VI.2) applied to $p, n - i \in \mathbb{Z}[i]$ (and the norm N as function g) finds $\pi \in \mathbb{Z}[i]$ with $(\pi) = (p, n-i)$. Here $p \in (\pi)$ implies $N(\pi)$ divides $N(p) = p^2$ and $n - i \in (\pi)$ implies that $N(\pi)$ divides $N(n-i) = n^2 + 1 < p^2$. As a result, $N(\pi) \in \{1, p\}$ and since $N(\pi) = 1$ would mean $\ker(\phi) = \mathbb{Z}[i]$ which is not the case, one concludes $N(\pi) = p$. This finishes the more constructive approach to writing such a prime p as a sum of two squares. It should be evident that the most time consuming step here is finding $x \in \mathbb{F}_p$ with $x^2 = -\overline{1}$. An efficient method (in practice) for this is given by the so-called *Tonelli-Shanks* algorithm.

VI.3.6 Example. Using the method described in Remark VI.3.5 we write the prime number $p = 7933$ as a sum of two squares. It turns out that $n = 2950$ satisfies $n^2 \equiv -1 \bmod p$. Hence it remains to find $\gcd(7933, 2950 - i)$ using the Euclidean algorithm in $\mathbb{Z}[i]$. Now $7933/(2950 - i) \approx 2.69$, and

$$7933 = 3 \cdot (2950 - i) - 917 + 3i.$$

Next, $(2950 - i)/(-917 + 3i) \approx -3.22$, and

$$2950 - i = -3 \cdot (-917 + 3i) + 199 + 8i.$$

One more division with remainder: $(-917 + 3i)/(199 + 8i) \approx -4.60 + 0.20i$, leading to

$$-917 + 3i = -5 \cdot (199 + 8i) + 78 + 43i.$$

Now $199+8i=(2-i)\cdot(78+43i)$, hence $\gcd(7933,2950-i)=78+43i$ and indeed

$$7933=78^2+43^2.$$

∎

VI.3.7 Corollary. *Suppose $n\in\mathbb{Z}_{>0}$ has prime factorization*

$$n=2^k p_1^{n_1}\dots p_r^{n_r} q_1^{m_1}\dots q_s^{m_s},\qquad n_j,m_j\in\mathbb{Z}_{>0}$$

with p_j, q_j pairwise distinct primes and $p_j\equiv 1 \bmod 4$ and $q_j\equiv 3 \bmod 4$.

Then n can be written as a sum of two squares precisely when:

$$m_j\equiv 0 \bmod 2,\qquad \text{for all}\quad j\in\{1,\dots s\}.$$

Proof. Observe that n can be written as a sum of two squares:

$$n=a^2+b^2=(a+bi)(a-bi)=\alpha\overline{\alpha}$$

precisely when some $\alpha\in\mathbb{Z}[i]$ exists with $\alpha\overline{\alpha}=n$.

Using Theorem VI.3.3 the prime factorization of n in $\mathbb{Z}[i]$, where we ignore the factorization of 2^k, is given by

$$n=2^k(\pi_1^{n_1}\overline{\pi}_1^{n_1})\dots(\pi_r^{n_r}\overline{\pi}_r^{n_r})q_1^{m_1}\dots q_s^{m_s}.$$

In case $n=\alpha\overline{\alpha}$ then the prime factorization of α is of the form

$$\alpha=u(1+i)^l(\pi_1^{a_1}\overline{\pi}_1^{b_1})\dots(\pi_r^{a_r}\overline{\pi}_r^{b_r})q_1^{c_1}\dots q_s^{c_s},$$

for a unit u. Hence (note that $u\bar{u}=1$):

$$n=\alpha\overline{\alpha}=2^l p_1^{a_1+b_1}\dots p_r^{a_r+b_r} q_1^{2c_1}\dots q_s^{2c_s}.$$

Since the prime factorization (here, of n) in $\mathbb{Z}[i]$ is unique,

$$m_j=2c_j,\qquad \text{so}\quad m_j\equiv 0 \bmod 2\quad \forall j.$$

Vice versa, suppose all m_j are even. We construct $\alpha\in\mathbb{Z}[i]$ with $\alpha\overline{\alpha}=n$ as follows. Put

$$c_j:=\frac{m_j}{2},\qquad l:=k.$$

Next, take

$$a_j\in\{0,1,\dots,n_j\}\qquad\text{and}\quad b_j:=n_j-a_j,$$

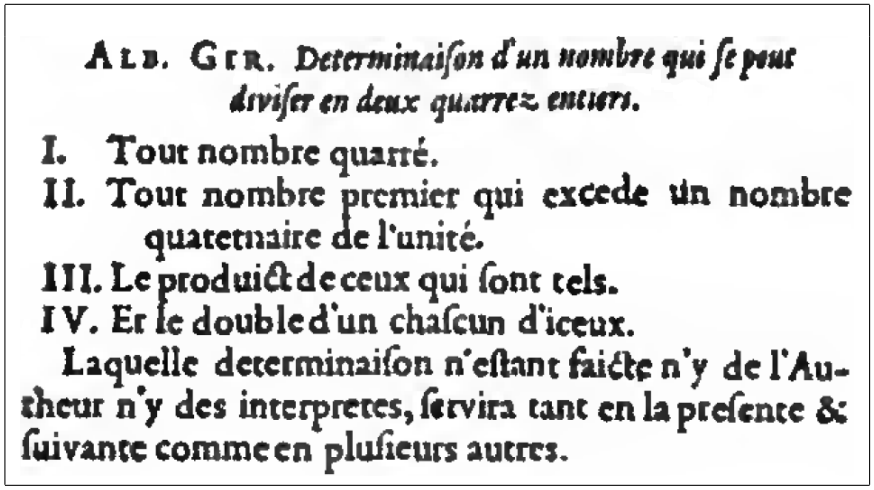
so b_j is determined by the choice of a_j. For u we take any of the 4 units in $\mathbb{Z}[i]$. This determines an α with the required properties.

We note that in this way one finds $4\cdot\prod_{i=1}^{r}(n_j+1)$ possible α's, which is exactly the number of elements in the set

$$\left\{(a,b)\in\mathbb{Z}^2\ :\ a^2+b^2=n\right\}.$$

This shows VI.3.7. ∎

VI.3.8 Remark. The result of Corollary VI.3.7 is usually attributed to Fermat, who stated it in a letter dated December 25th, 1640 he sent to the French priest Marin Mersenne. However, somewhat earlier the French-born mathematician and Albert Girard (1595–1632) who worked as an engineer in the army of the Dutch stadtholder Frederik Hendrik prince of Orange, edited and translated the works of Simon Stevin. In Chapter 5 of the first Volume (on Arithmetic), page 622 of the 1625 edition and page 156 of the edition published in 1634, one finds:



ALB. GIR. *Determinaiſon d'un nombre qui ſe peut diviſer en deux quarrez entiers.*

I. Tout nombre quarré.
II. Tout nombre premier qui excede un nombre quatemaire de l'unité.
III. Le produict de ceux qui ſont tels.
IV. Et le double d'un chaſcun d'iceux.

Laquelle determinaiſon n'eſtant faicte n'y de l'Autheur n'y des interpretes, ſervira tant en la preſente & ſuivante comme en pluſieurs autres.

So this presents, 15 years before Fermat's letter, the correct description of all natural numbers which are sums of two squares:

I The squares;
II The prime numbers $p \equiv 1 \bmod 4$;
III The products of numbers from I and II;
IV Twice any number from I, II, or III.

A complete *proof* of the assertions by Girard and Fermat came much later, presented by Euler and published in two consecutive issues of the journal *Novi commentarii Academiae Scientiarum Imperialis Petropolitanae* between 1752 and 1755.

VI.3.9 Example. Take $n = 41$, which is a prime congruent to 1 modulo 4 hence a sum of two squares. We have

$$41 = 16 + 25 = (4+5i)(4-5i) = \pi\overline{\pi},$$

with $\pi = 4+5i$ and $\overline{\pi}$ irreducible in $\mathbb{Z}[i]$.

For $n = 45$ one finds

$$45 = 5 \cdot 3^2 = (1+2i)(1-2i)3^2$$

with $1 \pm 2i$ and 3 irreducible in $\mathbb{Z}[i]$. Taking

$$\alpha = (1+2i)3 = 3+6i \qquad \text{one obtains} \quad 45 = \alpha\bar{\alpha} = 3^2 + 6^2.$$

Let $n = 65 = 5 \cdot 13$. As $5 = (1+2i)(1-2i)$ and $13 = (2+3i)(2-3i)$, the prime factorization of 65 in $\mathbb{Z}[i]$ is

$$65 = \pi_1\overline{\pi}_1\pi_2\overline{\pi}_2, \qquad \text{with} \quad \pi_1 = 1+2i, \quad \pi_2 = 2+3i.$$

Taking

$$\alpha = \pi_1\pi_2 \qquad \text{one finds} \quad \alpha = -4+7i \quad \text{and} \quad 65 = (-4)^2 + 7^2.$$

Taking

$$\alpha = \pi_1\bar{\pi}_2 \qquad \text{one has} \quad \alpha = 8+i \quad \text{and} \quad 65 = 8^2 + 1^2.$$

These are, up to signs and permutations of a, b, the only two representations of 65 as a sum of two squares. ∎

VI.4 Exercises

1. Let $\gamma = \frac{1}{2}(1+\sqrt{-19})$ and $R = \mathbb{Z}[\gamma] = \{a+b\gamma : a,b \in \mathbb{Z}\} \subset \mathbb{C}$. Define

$$N : R \longrightarrow \mathbb{Z}_{\geq 0}, \qquad N(a+b\gamma) := a^2 + ab + 5b^2.$$

   (a) Show that $\gamma^2 = \gamma - 5$ and that R is a subring of $\mathbb{C}$.

   (b) Show that $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in R$.

   (c) Let $\alpha \in R$. Prove that

$$\alpha \in R^\times \iff N(\alpha) = 1 \iff \alpha \in \{\pm 1\},$$

   hence $R^\times = \{1, -1\}$.

   (d) Prove that no surjective ring homomorphisms

$$\varphi : R \longrightarrow \mathbb{F}_2 \quad \text{or} \quad \varphi : R \longrightarrow \mathbb{F}_3$$

   exist (hint: use $\gamma^2 = \gamma - 5$ and use this to limit the possibilities for $\varphi(\gamma)$).

2. Take $R = \mathbb{Z}[\gamma]$ as in the previous exercise. Suppose that $g : R - \{0\} \to \mathbb{Z}_{\geq 0}$ satisfies the condition stated in Definition VI.1.1, and choose $b \in R - \{0, 1, -1\}$ with $g(b)$ as small as possible.

   (a) Show that b is not a unit and that

$$\forall a \in R : \ \exists r \in \{0, 1, -1\} : \quad a \equiv r \bmod Rb.$$

   (b) Prove that $R/Rb = \{\overline{0}, \overline{1}, \overline{-1}\}$, with $\overline{r} = (r + Rb)$. Conclude that $R/Rb \cong \mathbb{F}_2$ or $R/Rb \cong \mathbb{F}_3$.

   (c) Using Exercise 1 (d) above, deduce a contradiction.

   The conclusion of this problem: g as above does not exist, hence R is *not* Euclidean.

3. Let R and N be as in Exercise 1 above. For $a, b \in R$, $b \neq 0$ we say that division with remainder is possible for the pair (a, b) if $q, r \in R$ exist with

$$a = qb + r \quad \text{and} \quad N(r) < N(b).$$

   (a) Suppose that $(a, b) \in R \times R$ with $b \neq 0$, and that division with remainder is *not* possible for this pair. Prove that in this case division with remainder *is* possible for $(2a, b)$, as well as for one of the pairs $(\gamma a, b)$, $((1-\gamma)a, b)$. (Hint: draw a picture).

   (b) Show that $R2 + R\gamma = R$ and also $R2 + R(1-\gamma) = R$.

   (c) Prove that R is a principal ideal domain (imitate the proof of Theorem VI.1.4, but use (a) instead of the condition from Definition VI.1.1).

4. Calculate $\gcd(4+7i, 7-9i)$ in $\mathbb{Z}[i]$ and write $4+7i$ and $7-9i$ in $\mathbb{Z}[i]$ as a product of irreducible factors.

5. Let $p = 18313$, which is a prime number. It turns out that $n := 6731 \in \mathbb{Z}$ satisfies $n^2 \equiv -1 \bmod p$. Use this to write p as a sum of two squares (you will probably want to use a calculator for this).

6. Suppose $n = a^2 + b^2$. Find p, q in terms of a and b such that $2n = p^2 + q^2$. Similarly, find r, s with $5n = r^2 + s^2$.

7. Prove that for $m \in \{-2, 2, 3\}$ the ring $\mathbb{Z}[\sqrt{m}]$ is Euclidean, using $g(\alpha) = |N(\alpha)|$ (the norm).

8. Prove that for $m \in \{-11,-7,-3,5,13\}$ the ring $\mathbb{Z}[\frac{1}{2}(1+\sqrt{m})]$ is Euclidean, using

$$g(a+\frac{1}{2}(1+\sqrt{m})b) = |a^2+ab-\frac{m-1}{4}b^2|.$$

(Hint: first check that $g(x+y\sqrt{m}) = |(x+y\sqrt{m})(x-y\sqrt{m})|$ for $x, y \in \mathbb{Q}$.)

9. Put $R = \{a/b \in \mathbb{Q} : a, b \in \mathbb{Z},\ b \text{ odd }\}$, which is a subring of $\mathbb{Q}$.
   (a) Describe $R^\times$.
   (b) Show that every $x \in R$, $x \neq 0$ can be written in a unique way as $x = 2^k \cdot u$ with $k \in \mathbb{Z}_{\geq 0}$ and $u \in R^\times$.
   (c) Show that the function

$$g : R - \{0\} \longrightarrow \mathbb{Z}_{\geq 0}, \qquad g(x) = k \quad \text{if} \quad x = 2^k \cdot u$$

   as in (b), makes R a Euclidean ring.
   (d) Show that up to multiplication by units 2 is the only irreducible element of R. Is $2R$ a prime ideal of R?

10. The ring $R[[X]]$ of *formal power series* over a ring R consists of all expressions $\sum_{i=0}^{\infty} a_i X^i$ with $a_i \in R$. The addition and multiplication are the familiar ones for power series.
   (a) Suppose R is unitary and let $f = \sum_{i=0}^{\infty} a_i X^i \in R[[X]]$. Prove:

$$f \in R[[X]]^\times \iff a_0 \in R^\times.$$

   (b) Suppose that R is a *field*. Define

$$g : R[[X]] - \{0\} \longrightarrow \mathbb{Z}_{\geq 0}$$

   by

$$g(\sum_{i=0}^{\infty} a_i X^i) = \min\{i :\ a_i \neq 0\}.$$

   Prove that this function makes $R[[X]]$ a Euclidean ring.

11. Let R be a principal ideal domain and consider $a, b \in R$ with prime factorizations

$$a = u p_1^{n_1} \dots p_r^{n_r}, \qquad b = v p_1^{m_1} \dots p_r^{n_r}, \qquad n_i, m_i \in \mathbb{Z}_{\geq 0}.$$

Define $d \in R$ as in Definition V.4.5:

$$d := p_1^{k_1} \dots p_r^{k_r}, \qquad \text{with} \quad k_i := \min\{n_i, m_i\}.$$

Prove that $(a, b) = (d)$ (hint: examine the proof of Theorem V.3.5).

VII FIELDS

VII.1 Prime fields and characteristic

VII.1.1 Definition. Let K be a field. A subset $K' \subset K$ is called a *subfield* if the next three conditions hold:

- $1 \in K'$,
- $a, b \in K' \implies a - b \in K'$,
- $a, b \in K', b \neq 0 \implies ab^{-1} \in K'$.

In other words, a subset $K' \subseteq K$ of a field K is a subfield if by restricting the operations defined on K to K' it is itself a field. One easily verifies that the intersection of any collection of subfields of K is again a subfield.

VII.1.2 Definition. The intersection of *all* subfields of a field K is called the *prime field* K_0 of K, so

$$K_0 := \bigcap_{K' \subset K} K'$$

where the intersection is taken over all subfields $K' \subseteq K$.

By construction the prime field is the smallest (with respect to the inclusion) subfield of K. Note that $0, 1 \in K$ are in the prime field K_0.

VII.1.3 Theorem. *Let K be a field. The prime field of K is isomorphic to*

either the field $\mathbb{Q}$ of rational numbers,
or a field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, with p a prime number.

Proof. Let K_0 denote the prime field of K. Define

$$\kappa : \mathbb{Z} \longrightarrow K_0$$

by

$$\begin{array}{lcll} \kappa(n) & = & 1+1+\ldots+1 \in K_0 & (n \text{ terms}) \\ \kappa(0) & = & 0 \in K_0 & \\ \kappa(-n) & = & -(1+1+\ldots+1) \in K_0 & (n \text{ terms}) \end{array}$$

where $n \in \mathbb{Z}_{>0}$ and the 1 in the right hand side is $1 \in K$. Then κ is a ring homomorphism. One has $\kappa(\mathbb{Z}) \subset K_0$ since K_0 is a field and $1 \in K_0$.

Since K_0 is a field $\kappa(\mathbb{Z}) \subseteq K_0$ has no zero divisors. Moreover $\kappa(\mathbb{Z})$ contains $1 \neq 0$, hence $\kappa(\mathbb{Z})$ is a *domain*. By Theorem II.3.7 $\kappa(\mathbb{Z}) \cong \mathbb{Z}/\mathrm{Ker}(\kappa)$, so $\mathrm{Ker}(\kappa)$ is a prime ideal of $\mathbb{Z}$. We conclude $\mathrm{Ker}(\kappa) = \{0\}$ or $\mathrm{Ker}(\kappa) = p\mathbb{Z}$ for some prime number p.

First, assume $\mathrm{Ker}(\kappa) = 0$. In this case κ is injective and $\kappa(\mathbb{Z}) \cong \mathbb{Z}$. We extend κ to a map

$$\kappa_1 : \mathbb{Q} \to K_0, \qquad \kappa_1(a/b) := \kappa(a) \cdot (\kappa(b))^{-1}, \quad (a, b \in \mathbb{Z},\, b \neq 0).$$

This map is well defined and one easily verifies that $\kappa_1 : \mathbb{Q} \to K_0$ is a ring homomorphism. Since $\{0\}$ and $\mathbb{Q}$ are the only ideals of $\mathbb{Q}$ it follows that κ_1 is injective. Its image $\kappa_1(\mathbb{Q})$ is therefore a subfield of K_0. As K_0 is the *smallest* subfield of K, one concludes $K_0 = \kappa_1(\mathbb{Q}) \cong \mathbb{Q}$.

The remaining case is that $\mathrm{Ker}(\kappa) = p\mathbb{Z}$ for p prime. Then $\kappa(\mathbb{Z}) \cong \mathbb{Z}/p\mathbb{Z}$, which by Theorem I.2.11 is a field. So $\kappa(\mathbb{Z})$ is a subfield of K_0, which as above implies $\kappa(\mathbb{Z}) = K_0$. Hence $K_0 = \kappa(\mathbb{Z}) \cong \mathbb{Z}/p\mathbb{Z}$. ∎

VII.1.4 Definition. Let K be a field with prime field K_0.
If $K_0 \cong \mathbb{Q}$ we say that K has *characteristic zero*, notation: $\mathrm{char}(K) = 0$.
If $K_0 \cong \mathbb{F}_p$ we say that K has *characteristic* p, notation: $\mathrm{char}(K) = p$.

We see that in all cases $\mathrm{char}(K)$ is the non-negative generator of the ideal $\mathrm{Ker}(\kappa) \subset \mathbb{Z}$, with $\kappa : \mathbb{Z} \to K$ the unique ring homomorphism such that $\kappa(1) = 1$.

VII.2 Algebraic and transcendental

VII.2.1 Definition. If L is a field and $K \subset L$ is a subfield, then L is called an *extension field* or *field extension* of K.

If $L \supseteq K$ is an extension field of a field K and $\alpha \in L$, then we have the evaluation homomorphism (see Theorem III.2.1)

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

The kernel of ev_α is an ideal in $K[X]$, and therefore by Theorem III.4.1 it is a principal ideal. We have either $\mathrm{Ker}(\mathrm{ev}_\alpha) = (0)$ (equivalently, ev_α is injective) or $\mathrm{Ker}(\mathrm{ev}_\alpha) = (f)$ for some unique monic $f \in K[X]$.

VII.2.2 Definition. Given an extension of fields $K \subseteq L$ and $\alpha \in L$, one says that α is *transcendental over* K if the evaluation homomorphism $\mathrm{ev}_\alpha : K[X] \to L$ is injective.

We say that α is *algebraic over* K if a nonzero $f \in K[X]$ exists with $f(\alpha) = 0$. In this case, the unique monic generator of the ideal $\mathrm{Ker}(\mathrm{ev}_\alpha)$ is called the *minimal polynomial* of α over K. This minimal polynomial is denoted f_K^α.

If α is in an extension L of the field K, then

$$K[\alpha] := \mathrm{ev}_\alpha(K[X]) \cong K[X]/\mathrm{Ker}(\mathrm{ev}_\alpha)$$

is a subring $\neq (0)$ of the field L. Hence $K[\alpha]$ is a domain. In case α is transcendental over K then ev_α is injective hence $K[X] \cong K[\alpha] \subseteq L$. In case α is algebraic over K we have $K[X]/(f_K^\alpha) \cong K[\alpha] \subseteq L$. Since $K[\alpha]$ is a domain, Theorems IV.1.5 and V.2.4 and IV.2.3 show in this case that $f_K^\alpha \in K[X]$ is irreducible and that $K[\alpha]$ is a field. Note that for $g \in K[X]$ we have

$$\text{if } g(\alpha) = 0 \quad \text{then} \quad g \in \mathrm{Ker}(\mathrm{ev}_\alpha) = (f_K^\alpha) \Longrightarrow g = q f_K^\alpha \text{ for some } q \in K[X].$$

Regardless of α being algebraic or transcendental over K, the ring $K[\alpha]$ is a domain. This leads to the following.

VII.2.3 Notation. Let α be an element of an extension L of the field K. Then $K(\alpha)$ denotes the subfield of L given by

$$K(\alpha) := \{x \cdot y^{-1} \in L : \ x, y \in K[\alpha],\ y \neq 0\}.$$

In fact this is isomorphic to the field of fractions (see I.3) of the domain $K[\alpha]$.

VII.2.4 Definition. A field extension L of K is called *simple* if an $\alpha \in L$ exists with $L = K(\alpha)$.

We summarize the above in the next result.

VII.2.5 Theorem. *Let L be an extension of a field K and $\alpha \in L$ algebraic over K.*
Then the minimal polynomial f_K^α of α over K is irreducible in $K[X]$.
Moreover $K[\alpha] \cong K[X]/(f_K^\alpha)$ and $K[\alpha]$ is a field, indeed

$$K(\alpha) = K[\alpha].$$

Proof. By the first isomorphism theorem II.3.7 $\mathrm{ev}_\alpha : K[X] \to K[\alpha] \subset L$ yields an isomorphism $K[X]/\mathrm{Ker}(\mathrm{ev}_\alpha) \cong K[\alpha]$. Since L is a field, $K[\alpha]$ is a domain hence $\mathrm{Ker}(\mathrm{ev}_\alpha) = (f_K^\alpha)$ is a prime ideal in $K[X]$. Now $K[X]$ is a principal ideal domain so by Theorem V.2.4 f_K^α is irreducible. The same theorem implies that (f_K^α) is a maximal ideal, hence $K[\alpha]$ is a field. Since $K(\alpha)$ is the smallest subfield of L containing $K[\alpha]$, one concludes $K(\alpha) = K[\alpha]$. ∎

VII.2.6 Example. Take $d \in \mathbb{Q}$ such that $\alpha := \sqrt{d} \notin \mathbb{Q}$. Then $f_{\mathbb{Q}}^\alpha = X^2 - d$. This polynomial is irreducible since it has degree 2 and (by the choice of d) it has no zero in $\mathbb{Q}$. Moreover

$$\mathbb{Q}[\alpha] = \mathbb{Q}[\sqrt{d}\,] = \left\{a + b\sqrt{d} \in \mathbb{C} : \ a, b \in \mathbb{Q}\right\}.$$

We show that $\mathbb{Q}[\sqrt{d}]$ is indeed a field by presenting the inverse of $x = a + b\sqrt{d} \neq 0$. Define

$$\overline{x} := a - b\sqrt{d} \quad \text{and} \quad N(x) = x\overline{x} = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - b^2 d \quad (\in \mathbb{Q}).$$

Since $d \in \mathbb{Q}$ is not a square and $x \neq 0$, it follows that $N(x) \neq 0$ (check for yourself!). Hence

$$x^{-1} := \frac{\overline{x}}{N(x)} = \frac{a}{a^2 - b^2 d} - \frac{b}{a^2 - b^2 d}\sqrt{d} \quad \in \mathbb{Q}[\sqrt{d}].$$

—∎

Note that in case α is transcendental over K, then $K(\alpha) \cong K(X)$, the field of rational functions in the variable X and coefficients in K. Moreover $K[\alpha] \cong K[X]$. In particular $\alpha^{-1} \notin K[\alpha]$ and $K(\alpha) \supsetneq K[\alpha]$.

On the other hand, if α is algebraic over K then $K[\alpha] = K(\alpha)$ hence $\alpha^{-1} \in K[\alpha]$ whenever $\alpha \neq 0$.

VII.2.7 Remark. We note that the difference transcendental/algebraic in this section is analogous to the difference characteristic zero/characteristic $p > 0$ in VII.1.4. In this analogy the evaluation homomorphism

$$\mathrm{ev}_\alpha : K[X] \to K[\alpha], \quad \mathrm{ev}_\alpha(X) := \alpha$$

corresponds to the ring homomorphism

$$\kappa : \mathbb{Z} \to K, \quad \kappa(1) := 1$$

used in the proof of VII.1.3. Also, f_K^α corresponds to p.

VII.2.8 Example. Taking $L = K(X)$ and $\alpha = X \in K(X)$ we obtain an easy example of a simple extension L of K with a transcendental α.

Much less evident is the existence of numbers in $\mathbb{R}$ or in $\mathbb{C}$ which are transcendental over $\mathbb{Q}$. Such numbers are called *transcendental numbers* (without mentioning any fields). So a transcendental number is not a zero of any polynomial ($\neq 0$) with coefficients in $\mathbb{Q}$. By a counting argument (see Exercise 2 on page 100) one can show the existence of transcendental numbers. The first person who found an explicit one was Liouville (Joseph Liouville, French mathematician, 1809-1882): he showed that $\sum_{k=1}^{\infty} 10^{-k!}$ is transcendental. In 1933 the English statistician and economist D.G. Champernowne (1912–2000) while he was still an undergraduate student published a paper showing that $0,12345678910111213\ldots$ is *normal* (which means that each of the digits $0, 1, \ldots, 9$ occurs "in the limit" equally often, and similarly for each pair $01, 02, \ldots, 98, 99$ et cetera). In 1937 Kurt Mahler (Jewish-German and later Australian mathematician, 1903–1988) who worked in Groningen between 1934 and 1936 published the result that in fact Champernowne's constant is transcendental. In 1873 Hermite (Charles Hermite, French mathematician, 1822–1901) showed that $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ is transcendental, and in 1882 Lindemann (Carl Louis Ferdinand von Lindemann, German mathematician) did the same for $\pi = 3.14159\cdots$. For more about this, see for example the textbook I. Stewart, *Galois Theory*, Chapter 6. ∎

VII.2.9 Example. The complex number $i \in \mathbb{C}$ is algebraic over $\mathbb{R}$, with $f_{\mathbb{R}}^{i} := X^2 + 1$. It is even algebraic over $\mathbb{Q}$ with $f_{\mathbb{Q}}^{i} = X^2 + 1$. If $a \in \mathbb{R}$ is transcendental then $ia \in \mathbb{C}$ is algebraic over $\mathbb{R}$, with $f_{\mathbb{R}}^{ia} = X^2 + a^2$. However ia is not algebraic over $\mathbb{Q}$ (why?).

For any $\alpha \in L$ one has $f_{L}^{\alpha} = X - \alpha$. The minimal polynomial of α over a subfield K of$L$ will in general depend on the choice of K.

For all $k, n \in \mathbb{Z}_{>0}$ the number $\alpha = \sqrt[n]{k} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ since it is a zero of $X^n - k \in \mathbb{Q}[X]$. The complex numbers

$$e^{\frac{2\pi i k}{n}} := \cos\frac{2\pi k}{n} + i \sin\frac{2\pi k}{n} \qquad (k \in \mathbb{Z})$$

are also algebraic over $\mathbb{Q}$: they are zero's of $X^n - 1$. These latter numbers are called *roots of unity*.

The problem of finding the minimal polynomial of an algebraic α over a given field K will be addressed in Section VII.4. ∎

VII.2.10 Example. Let $\alpha \in \mathbb{C} = L$ be a zero of $f = X^3 + X^2 - 1$. Then α is algebraic over $\mathbb{Q} = K$ (since $f \in \mathrm{Ker}(\mathrm{ev}_\alpha)$). The polynomial f is irreducible in $\mathbb{Q}[X]$ since $\deg(f) = 3$ and f has no zero in $\mathbb{Z}$ and therefore not in $\mathbb{Q}$ (see V.4.11). Hence $f_{\mathbb{Q}}^{\alpha} = f$ and $\mathbb{Q}[\alpha] \cong \mathbb{Q}[X]/(f)$ is a field. We present a method for finding the inverse of $a \in \mathbb{Q}[\alpha]$ with $a \neq 0$. (A different method which uses the Euclidean algorithm is given in VI.2.5).

From Theorem III.3.4 we know that every $a \in \mathbb{Q}[\alpha]$ an be written in a unique way as

$$a = a_0 + a_1\alpha + a_2\alpha^2 \qquad \text{with} \quad a_0, a_1 a_2 \in \mathbb{Q}.$$

To determine the inverse we have to solve the equation

$$ax = 1 \qquad \text{i.e.,} \quad (a_0 + a_1\alpha + a_2\alpha^2)(x_0 + x_1\alpha + x_2\alpha^2) = 1$$

for $x_i \in \mathbb{Q}$. Since $1, \alpha, \alpha^2$ are linearly independent over $\mathbb{Q}$ (as follows from the unicity of the presentation $a = a_0 + a_1\alpha + a_2\alpha^2$) and

$$\alpha^3 = -\alpha^2 + 1, \qquad \alpha^4 = \alpha \cdot \alpha^3 = -\alpha^3 + \alpha = \alpha^2 + \alpha - 1,$$

we need to solve (for given a_i) the system

$$\begin{cases} a_0x_0+ & a_2x_1+ & (a_1-a_2)x_2 & =1 \\ a_1x_0+ & a_0x_1+ & a_2x_2 & =0 \\ a_2x_0+ & (a_1-a_2)x_1+ & (a_0-a_1+a_2)x_2 & =0 \end{cases}$$

In matrix notation this can be written as

$$\begin{pmatrix} a_0 & a_2 & a_1-a_2 \\ a_1 & a_0 & a_2 \\ a_2 & a_1-a_2 & a_0-a_1+a_2 \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The problem is therefore solved if we invert the 3×3 matrix M given here.

In the special case

$$a=1+\alpha^2, \qquad \text{so} \quad (a_0,a_1,a_2)=(1,0,1),$$

the inverse is easily found using Gauss-elimination: in this case

$$M=\begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & -1 & 2 \end{pmatrix} \quad \text{and} \quad M^{-1}=\frac{1}{5}\begin{pmatrix} 3 & -1 & 2 \\ 1 & 3 & -1 \\ -1 & 2 & 1 \end{pmatrix}.$$

Now $M^{-1}(1,0,0)=\frac{1}{5}(3,1,-1)$ and therefore

$$a=1+\alpha^2 \quad \Longrightarrow \quad \alpha^{-1}=\frac{3}{5}+\frac{1}{5}\alpha-\frac{1}{5}\alpha^2.$$

VII.3 Finite and algebraic extensions

Suppose L is an extension of a field K. Then L can be considered as a *vector space* over K. This means that one considers elements of K as *scalars* (with the usual properties of a field), and the elements of L are considered as *vectors* (forming an additive group). Vectors are multiplied by scalars using that K is a subfield of L.

In this way (for example) the complex numbers are considered as vectors over $\mathbb{R}$. Using the real and imaginary part of a complex number there is even an isomorphism of vector spaces over $\mathbb{R}$: $\mathbb{C}\cong\mathbb{R}^2$.

VII.3.1 Definition. Let L be an extension of a field K. We say that L is *finite* over K if the dimension of L considered as a vector space over K is finite.

The *degree* of L over K, notation: $[L:K]$, is the dimension of L considered as K-vector space:

$$[L:K]:=\dim_K(L).$$

We call L *algebraic* over K if *every* $\alpha\in L$ is algebraic over K (see VII.2.2).

VII.3.2 Example. $[\mathbb{C}:\mathbb{R}]=2$ and $[\mathbb{Q}(\sqrt[3]{2}):\mathbb{Q}]=3$. Namely, the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is X^3-2. Therefore $\mathbb{Q}(\sqrt[3]{2})\cong\mathbb{Q}[X]/(X^3-2)$ and Theorem III.3.4 implies that this is a vector space of dimension 3 over $\mathbb{Q}$. See Theorem VII.3.3 below for a generalisation of this argument. Since every finite dimensional vector space over $\mathbb{Q}$ is countable (if S and T are both countable, so is $S\times T$) we conclude that $\mathbb{R}$ is *not* finite over $\mathbb{Q}$.

VII.3.3 Theorem. *Let L be an extension of a field K and $\alpha \in L$. Then*

$$\alpha \text{ is algebraic over } K \iff K(\alpha) \text{ is finite over } K.$$

Moreover if α is algebraic over K then

$$n := [K(\alpha) : K] = \deg(f_K^\alpha) \qquad \text{and} \quad 1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$$

is a K-basis of the vector space $K(\alpha)$ over K.

Proof. $\Leftarrow$: Suppose $[K(\alpha) : K] = n < \infty$. Since every $(n+1)$-tuple of vectors in an n-dimensional vector space is linearly dependent, a linear relation

$$a_0 \cdot 1 + a_1 \cdot \alpha + \ldots + a_n \cdot \alpha^n = 0$$

with $a_0, a_1, \ldots, a_n \in K$ not all zero exists between $1, \alpha, \alpha^2, \ldots, \alpha^n \in K(\alpha)$. Hence α is a zero of the polynomial $a_0 + a_1 X + \ldots + a_n X^n \in K[X]$, so α is algebraic over K.
$\Rightarrow$: Take α algebraic over K. By Theorem VII.2.5 $K(\alpha) = K[\alpha] \cong K[X]/(f_K^\alpha)$. Next, using Theorem III.3.4 every element of the latter ring is represented uniquely as $a_0 + a_1 X + \ldots + a_{n-1} X^{n-1} + (f_K^\alpha)$ with $a_i \in K$ and $n := \deg(f_K^\alpha)$. Hence the classes of $1, X, \ldots, X^{n-1}$ are a basis of the linear space $K[X]/(f_K^\alpha)$ over K.

The isomorphism $K[X]/(f_K^\alpha) \cong K[\alpha]$ induced by ev_α sends $X^i + (f_K^\alpha)$ to α^i and it is K-linear. Hence $n = \deg(f_K^\alpha) = \dim_K(K[X]/(f_K^\alpha)) = \dim_K(K[\alpha])$ and $1, \alpha, \ldots, \alpha^{n-1}$ is a K-basis of $K[\alpha] = K(\alpha)$. This shows VII.3.3. ∎

VII.3.4 Theorem. *If $L \supset K$ is a finite extension of fields then L is algebraic over K.*

Proof. Take $\alpha \in L$. Since L is finite over K and $K(\alpha) \subseteq L$ is a K-linear subspace, $K(\alpha)$ is finite over K as well. Then VII.3.3 implies that α is algebraic over K. As $\alpha \in L$ is taken arbitrarily one concludes that L is algebraic over K. This proves VII.3.4. ∎

VII.3.3 and VII.3.4 imply immediately: if α is algebraic over K then $K(\alpha)$ is algebraic over K, so every $\beta \in K(\alpha)$ is algebraic over K. At the end of this chapter we will indicate how one may find the minimal polynomial of such β over K, see VII.4. As Exercise 9 on page 100 will show, the converse of VII.3.4 does not hold in general: there exist fields K which admit an algebraic extension that is not finite over K.

VII.3.5 Theorem. *Let K be a field and L an extension of K and M an extension of L (so $K \subseteq L \subseteq M$). Then*

$$M \text{ is finite over } K \iff M \text{ is finite over } L \text{ and } L \text{ is finite over } K.$$

Moreover if M is finite over K then

$$[M : K] = [M : L] \cdot [L : K].$$

Proof. $\Rightarrow$: Suppose that M is finite over K. Since L is a K-linear subspace of the K-vector space M, one concludes that L is finite over K. Take $\alpha_1, \ldots, \alpha_n \in M$ which span the vector space M over K. Every $x \in M$ can be expressed as $\sum_{i=1}^n a_i \alpha_i$ with $a_i \in K$. Then also $a_i \in L$ hence over L the vector space M is spanned by $\alpha_1, \ldots, \alpha_n$, and therefore $[M : L] \leq n$ which shows that M is finite over L.
$\Leftarrow$: Suppose that $[M : L] = n$ and $[L : K] = m$. Choose bases $\alpha_1, \alpha_2, \ldots, \alpha_m$ of L over K and $\beta_1, \beta_2, \ldots, \beta_n$ of M over L. We claim that $\{\alpha_i \beta_j : 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis of M over K.

Indeed, every $x \in M$ can be written as

$$x = \sum_{j=1}^{n} y_j \beta_j \quad \text{with } y_1, \ldots, y_n \in L.$$

As $\alpha_1, \ldots, \alpha_m$ is a basis of L over K, every y_j can be written as

$$y_j = \sum_{i=1}^{m} a_{ij} \alpha_i \quad \text{with } a_{ij} \in K \quad (1 \le i \le m, 1 \le j \le n).$$

We find

$$x = \sum_{1 \le i \le m, 1 \le j \le n} a_{ij} \alpha_i \beta_j.$$

This shows that any $x \in M$ is a K-linear combination of the $\alpha_i \beta_j$'s, as asserted by the claim.

It remains to show that $\{\alpha_i \beta_j : 1 \le i \le m, 1 \le j \le n\}$ is a *basis* of M over K. For this we must verify that they are linearly independent. Suppose

$$\sum_{1 \le i \le m, 1 \le j \le n} c_{ij} \alpha_i \beta_j = 0, \quad \text{with } c_{ij} \in K.$$

Then

$$\sum_{j=1}^{n} \Big(\sum_{i=1}^{m} c_{ij} \alpha_i \Big) \beta_j = 0 \quad \text{with } \sum_{i=1}^{m} c_{ij} \alpha_i \in L.$$

The β_j's are linearly independent over L, hence

$$\sum_{i=1}^{m} c_{ij} \alpha_i = 0$$

for $j = 1, 2, \ldots, n$. As the α_i's are linearly independent over K, one concludes

$$c_{ij} = 0$$

for all i and j. So indeed the $\alpha_i \beta_j$'s are linearly independent over K.

This implies that $\dim_K(M) = mn < \infty$ and

$$[M : K] = m \cdot n = [L : K] \cdot [M : L].$$

The proof of VII.3.5 is now complete. ∎

VII.3.6 Notation. If L is an extension of a field K and $\alpha_1, \alpha_2, \ldots, \alpha_n \in L$, then one defines inductively

$$K(\alpha_1, \alpha_2, \ldots, \alpha_n) := K(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})(\alpha_n).$$

VII.3.7 Corollary. *Let L be an extension of a field K and suppose $\alpha_1, \alpha_2, \ldots, \alpha_n \in L$ are algebraic over K. Then $K(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is finite over K.*

Proof. Use induction with respect to n. For $n = 1$ one applies Theorem VII.3.3. If $n > 1$ then put $K' = K(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$. The induction hypothesis implies that K' is *finite* over K. As α_n is algebraic over K, it is certainly algebraic over K'. Hence $K'(\alpha_n)$ is finite over K'. Theorem VII.3.5 (with $L = K', M = K'(\alpha_n)$) now implies that $K'(\alpha_n) = K(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is finite over K, as desired. This shows VII.3.7. ∎

VII.3.8 Theorem. *Let L be an extension of a field K, then*

(i) *if $\alpha, \beta \in L$ are algebraic over K then also*

$$\alpha + \beta, \quad \alpha - \beta, \quad \alpha\beta, \quad \alpha/\beta \quad (\beta \neq 0)$$

are algebraic over K.

(ii) *the set* $\{\alpha \in L : \alpha \text{ is algebraic over } K\}$ *is a subfield of* L *containing* K.

Proof. (i) By VII.3.7 $K(\alpha,\beta)$ is finite over K, hence algebraic (see VII.3.4). Definition VII.3.1 now implies that all elements of $K(\alpha,\beta)$, in particular $\alpha \pm \beta$, $\alpha\beta$, and α/β ($\beta \neq 0$) are algebraic over K.

(ii) From (i) and the definition of a subfield it follows that the given set is a subfield M of L. Clearly every $\alpha \in K$ is algebraic over K, hence $K \subseteq M$. This finishes the proof of VII.3.8.

■

VII.3.9 Definition. If L is an extension of a field K then the field

$$\{\alpha \in L \ : \ \alpha \text{ is algebraic over } K\}$$

discussed in Theorem VII.3.8 (ii) is called the *algebraic closure of* K *in* L.

The final theorem of this section is the analog of VII.3.5, with 'finite' replaced by 'algebraic'.

VII.3.10 Theorem. *If* $K \subseteq L \subseteq M$ *are fields then*

$$M \text{ is algebraic over } K \iff M \text{ is algebraic over } L \text{ and } L \text{ is algebraic over } K.$$

Proof. $\Rightarrow$: This is immediate from the definitions, as the reader should verify as an exercise.
$\Leftarrow$: Suppose M is algebraic over L and L is algebraic over K. Let $\alpha \in M$. We must show that α is algebraic over K. As M is algebraic over L, there are $n \in \mathbb{Z}_{>0}$ and $\beta_1, \beta_2, \ldots, \beta_n \in L$ with

$$\alpha^n + \beta_1\alpha^{n-1} + \ldots + \beta_{n-1}\alpha + \beta_n = 0.$$

This shows that α is algebraic over the subfield $K' = K(\beta_1, \beta_2, \ldots, \beta_n)$ of L, hence $K'(\alpha)$ is finite over K'. Since L is algebraic over K, all β_i are algebraic over K, hence by VII.3.7 $K' = K(\beta_1, \beta_2, \ldots, \beta_n)$ is finite over K. Now applying VII.3.5 to $K \subset K' \subset K'(\alpha)$ shows that $K'(\alpha)$ is finite over K. Then by VII.3.4 $K'(\alpha)$ is algebraic over K, and in particular α is algebraic over K, as desired. This shows VII.3.10. ■

VII.4 Determining a minimal polynomial

Suppose L is a finite extension of a field K and $\beta \in L$. How can the minimal polynomial f_K^β be found? (Theorem VII.3.4 shows it *exists*.) We will illustrate three methods for this in the special case $K = \mathbb{Q}, L = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ (compare Exercise 11 on page 100) and $\beta = 1 + \sqrt{2} + \sqrt{3}$.

1. The first method uses techniques from Linear Algebra. We choose a basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$, for example $1, \sqrt{2}, \sqrt{3}, \sqrt{2}\cdot\sqrt{3}$ (see the proof of VII.3.5 and also Exercise 11 on page 100). Using this basis we express elements of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ as vectors: the vector (a,b,c,d) represents the element $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{2}\cdot\sqrt{3}$. Now write powers $\beta^0, \beta^1, \beta^2, \ldots$ of β as such vectors:

$$\begin{array}{lll} \beta^0 = & 1 & = (1,0,0,0) \\ \beta^1 = & \beta & = (1,1,1,0) \\ \beta^2 = & & = (6,2,2,2) \\ \beta^3 = & & = (16,14,12,6) \\ \beta^4 = & & = (80,48,40,32). \end{array}$$

Continue until the vectors obtained in this way are linearly dependent. In the present case this happens for the first time when β^4 is considered. Using techniques for solving systems of linear equations one finds the linear relation

$$\beta^4 - 4\beta^3 - 4\beta^2 + 16\beta - 8 = 0.$$

Hence $f_{\mathbb{Q}}^{\beta} = X^4 - 4X^3 - 4X^2 + 16X - 8$, since a relation of lower degree would mean that the vectors corresponding to $\beta^0, \beta^1, \beta^2, \beta^3$ were linearly dependent, which they are not.
In finding out how many powers $\beta^0, \beta^1, \ldots$ are necessary until the corresponding vectors will be dependent, Exercise 12 on page 12 is useful.

2. The second method we now discuss relies on ideas from the field of '*Galois theory*'. It argues that since $f_{\mathbb{Q}}^{\sqrt{2}} = X^2 - 2$ has $\pm\sqrt{2}$ as zero's and similarly $f_{\mathbb{Q}}^{\sqrt{3}}$ has zero's $\pm\sqrt{3}$, it is natural to assume that $f_{\mathbb{Q}}^{1+\sqrt{2}+\sqrt{3}}$ will have the four numbers $1 \pm \sqrt{2} \pm \sqrt{3}$ as zero's. Computing the monic degree 4 polynomial

$$(X - (1 + \sqrt{2} + \sqrt{3})) \cdot (X - (1 + \sqrt{2} - \sqrt{3})) \cdot (X - (1 - \sqrt{2} + \sqrt{3})) \cdot (X - (1 - \sqrt{2} - \sqrt{3}))$$

one obtains

$$X^4 - 4X^3 - 4X^2 + 16X - 8$$

which has *rational* coefficients and by construction $1 + \sqrt{2} + \sqrt{3}$ as zero. To show that this polynomial is indeed $f_{\mathbb{Q}}^{\beta}$ it suffices to show that it is irreducible in $\mathbb{Q}[X]$.

3. The third method may be described as 'skillful computing': try to get rid of the square roots appearing in $\beta = 1 + \sqrt{2} + \sqrt{3}$. This may be done as follows:

$$\begin{aligned} \beta - 1 &= \sqrt{2} + \sqrt{3}, \text{ so} \\ (\beta - 1)^2 &= (\sqrt{2} + \sqrt{3})^2 = 2 + 2\sqrt{2} \cdot \sqrt{3} + 3 \\ &= 5 + 2\sqrt{2} \cdot \sqrt{3}, \text{ so} \\ ((\beta - 1)^2 - 5)^2 &= (2\sqrt{2} \cdot \sqrt{3})^2 = 24. \end{aligned}$$

Rewriting this final relation shows again that β is a zero of

$$((X - 1)^2 - 5)^2 - 24 = X^4 - 4X^3 - 4X^2 + 16X - 8.$$

To show that this polynomial is irreducible over $\mathbb{Q}$ it suffices to verify that the minimal polynomial of β has degree 4 over $\mathbb{Q}$. By VII.3.3 this is the same as showing $[\mathbb{Q}(\beta) : \mathbb{Q}] = 4$. Indeed: the computation above shows $\sqrt{2} \cdot \sqrt{3} \in \mathbb{Q}(\beta)$, hence also $(\beta - 1)\sqrt{2}\sqrt{3} = 2\sqrt{3} + 3\sqrt{2} \in \mathbb{Q}(\beta)$. Taking appropriate linear combinations of $2\sqrt{3} + 3\sqrt{2}$ and $\beta - 1 = \sqrt{2} + \sqrt{3}$ one finds that $\sqrt{2}, \sqrt{3} \in \mathbb{Q}(\beta)$, and therefore all of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is contained in $\mathbb{Q}(\beta)$. Evidently $\mathbb{Q}(\beta) \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$ as well, so $\mathbb{Q}(\beta) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. By Exercise 11 on page 100 this field has degree 4 over $\mathbb{Q}$, as desired.

VII.5 Exercises

1. Show that every $\alpha \in \mathbb{Q}(\sqrt{2})$ is algebraic over $\mathbb{Q}$ is.
2. Prove that the set $\{\alpha \in \mathbb{C} : \ \alpha \text{ is algebraic over } \mathbb{Q}\}$ is countable (hint: $\mathbb{Q}[X]$ is countable and every $f \in \mathbb{Q}[X]$, $f \neq 0$ has only finitely many zeros in $\mathbb{C}$). Conclude that complex (and even real) numbers exist that are transcendental over $\mathbb{Q}$.
3. Does $\alpha \in \mathbb{R}$ exist with $\mathbb{Q}(\alpha) = \mathbb{R}$? (hint: consider cardinalities.)
4. Show that $f_{\mathbb{Q}}^{\sqrt[n]{2}} = X^n - 2$ for every $n \in \mathbb{Z}_{>1}$.
5. Let α be algebraic over a field K and $f_K^\alpha = \sum_{i=0}^n a_i X^i$ of degree n. Show: if $\alpha \neq 0$ then $a_0 \neq 0$ and $\alpha^{-1} = \sum_{i=1}^n -a_0^{-1} a_i \alpha^{i-1}$.
6. Determine $f_{\mathbb{Q}}^\alpha$ and $\dim_{\mathbb{Q}} \mathbb{Q}(\alpha)$ for each of the following α's:

   $2-\sqrt{3}$, $\sqrt[3]{2}+\sqrt[3]{4}$, $\sqrt{3+2\sqrt{2}}$; β^{-1}, $\beta+1$ with β satisfying $\beta^3+3\beta-3=0$.
7. (a) Show that $\mathbb{Q}(\sqrt{2})(\sqrt{7}) = \mathbb{Q}(\sqrt{2}+\sqrt{7})$.

   (b) Prove: $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}+\sqrt{7}) = 4$.

   (c) Determine $f_{\mathbb{Q}}^{\sqrt{2}+\sqrt{7}}$.
8. Take $\alpha \in \mathbb{R}$ with $\alpha^3 - \alpha - 1 = 0$. Write each of the following elements in the form $a + b\alpha + c\alpha^2$ with $a, b, c \in \mathbb{Q}$:

$$\alpha^{10}, \quad \alpha^{-10}, \quad (\alpha^2+\alpha+1)^2, \quad (\alpha^2+1)^{-1}.$$

9. Take $L = \cup_{n=1}^\infty \mathbb{Q}(\sqrt[n]{2})$. Prove that:

   (a) L is a field (first show $\mathbb{Q}(\sqrt[n]{2}) \cup \mathbb{Q}(\sqrt[m]{2}) \subset \mathbb{Q}(\sqrt[nm]{2})$);

   (b) L is algebraic over $\mathbb{Q}$;

   (c) for every $n \in \mathbb{Z}_{\geq 1}$ the field L contains a subfield of degree n over $\mathbb{Q}$;

   (d) L is *not* finite over $\mathbb{Q}$.
10. Show that if α, β in some extension of a field K are algebraic over K, then

$$[K(\alpha,\beta) : K] \leq [K(\alpha) : K] \cdot [K(\beta) : K].$$

11. (a) Show that no $a, b \in \mathbb{Q}$ exist with $(a + b\sqrt{2})^2 = 3$, and conclude that $X^2 - 3$ is irreducible in $\mathbb{Q}(\sqrt{2})[X]$.

    (b) Prove: $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$.
12. Let L be a *finite* extension of a field K and $\alpha \in L$. Show that $\deg(f_K^\alpha)$ is a *divisor* of $[L : K]$.
13. Let $f = X^4 - 4X^3 - 4X^2 + 16X - 8$. Show that $\frac{1}{8} \cdot X^4 f(2/X)$ is an Eisenstein polynomial in $\mathbb{Z}[X]$. Conclude that f is irreducible in $\mathbb{Q}[X]$.
14. Let $\beta = 1 + \sqrt{2} + \sqrt{3}$. Express $\sqrt{2}$, $\sqrt{3}$, and β^{-1} as $\mathbb{Q}$-linear combinations of the $\mathbb{Q}$-basis $1, \beta, \beta^2, \beta^3$ of $\mathbb{Q}(\beta)$.
15. (a) Prove: $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) = \mathbb{Q}(\sqrt{2} \cdot \sqrt[3]{5}) = \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$.

    (b) Determine $f_{\mathbb{Q}}^\alpha$ for $\alpha = \sqrt{2} \cdot \sqrt[3]{5}$ and for $\alpha = \sqrt{2} + \sqrt[3]{5}$.
16. (a) Verify that $X^5 - 1 = (X-1)(X^4 + X^3 + X^2 + X + 1) =: (X-1)\Phi_5$ and that Φ_5 is irreducible in $\mathbb{Q}[X]$ (hint: use $\mathrm{ev}_{X+1} : \mathbb{Q}[X] \to \mathbb{Q}[X]$).

    (b) Let

$$M := \mathbb{Q}[X]/(\Phi_5), \quad \zeta := X + (\Phi_5), \quad \beta := X + X^4 + (\Phi_5) \in M, \quad L := \mathbb{Q}[\beta] \subset M.$$

    Find $a, b \in \mathbb{Q}$ with $\beta^2 = a\beta + b$ and determine $f_{\mathbb{Q}}^\beta$.

    (c) Find $[M : L]$ and f_L^ζ.

(d) Find an expression for $\cos\frac{2\pi}{5}$ containing only rational numbers and square roots of rational numbers.

17. Let $\alpha \in \mathbb{R}$, $\alpha^3 - \alpha - 1 = 0$. For each of the following elements determine the minimal polynomial over $\mathbb{Q}$:

$$\alpha - 1, \quad \alpha^2 + \alpha + 1, \quad (\alpha^2 + 1)^{-1}.$$

18. Let α be algebraic over a field K and suppose that $[K(\alpha) : K]$ is *odd*. Show that $K(\alpha) = K(\alpha^2)$.
19. Let L be an extension of a field K and let $\alpha, \beta \in L$ be algebraic over K. Suppose that $[K(\alpha) : K]$ and $[K(\beta) : K]$ are *coprime*.
Prove: $[K(\alpha, \beta) : K] = [K(\alpha) : K] \cdot [K(\beta) : K]$.
20. Let L be an extension of a field K and let K_0 be the algebraic closure of K in L (see VII.3.9).
Prove: every $\alpha \in L, \alpha \notin K_0$ is transcendental over K_0.
21. Let α be transcendental over a field K and $\beta \in K(\alpha)$, $\beta \notin K$. Prove:
    (a) α is algebraic over $K(\beta)$ (hint: let $\beta = f(\alpha)/g(\alpha)$ and consider the polynomial $f(X) - \beta g(X) \in K(\beta)[X]$).
    (b) β is transcendental over K.
22. Let K be a field.
    (a) ('splitting fractions'). Prove that the following collection is a basis of $K(X)$ over K:
    $$\{X^n : n \in \mathbb{Z}_{\geq 0}\} \cup \{X^i \cdot f^{-m} : f \in K[X]\},$$
    with in the second set only irreducible monic $f \in K[X]$ and $m \in \mathbb{Z}_{>0}, 0 \leq i < \deg(f)\}$.
    (b) Let α be transcendental over K. Prove that $[K(\alpha) : K]$ equals the cardinality of K in case K is infinite, and $[K(\alpha) : K]$ is countable infinite in case K is finite.
23. Take $K = \mathbb{F}_2(X, Y) = Q(\mathbb{F}_2[X, Y])$, (the field of fractions of $\mathbb{F}_2[X, Y]$).
    (a) Let $f = T^2 + X \in K[T]$. Prove that f is irreducible. Then consider
    $$L := K[T]/(f), \qquad t := T + (f) \ \in L.$$
    (b) Put $g = S^2 + Y \in L[S]$. Prove that g is irreducible. Next, take
    $$M := L[S]/(g), \qquad s := S + (g) \ \in M.$$
    (c) Observe that $K \subset L \subset M$ and show that 1, t, s, st is a K-basis of M.
    (d) Prove that every $\alpha \in M$, $\alpha \notin K$ satisfies $\deg(f_K^\alpha) = 2$. Conclude that the extension $M \supset K$ is not simple.

VIII AUTOMORPHISMS OF FIELDS AND SPLITTING FIELDS

VIII.1 Homomorphisms of fields

Recall (see Definition II.1.1) that if K and L are fields then $\phi : K \to L$ is called a *homomorphism of fields* (also called 'field homomorphism') if ϕ is a unitary ring homomorphism. In particular it satisfies $\phi(1) = 1$.

A homomorphism of fields has the property

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

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

Using $\phi(1 + 1 + \ldots + 1) = 1 + 1 + \ldots + 1$ it follows that the image of the prime field K_0 of K (see VII.1.1) is the prime field L_0 of L. Hence ϕ yields an isomorphism from K_0 to L_0. In case $K \subset L$ then $K_0 = L_0$ and the restriction of ϕ to K_0 is the identity map.

The image $\phi(K)$ of a field homomorphism $\phi : K \to L$ is also a field. Every field homomorphism is injective, since the only ideal in K are (0) and K and $1 \notin \mathrm{Ker}(\phi)$. A field homomorphism need not be surjective. For example the inclusion $\mathbb{R} \hookrightarrow \mathbb{C}$ is not. Even if $K = L$ a field homomorphism needs not be surjective, as illustrated by Example VIII.1.2.

The composition of field homomorphisms

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

is also a field homomorphism, as one easily verifies.

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

VIII.1.1 Theorem. *Let K be a field such that* $\mathrm{char}(K) = p > 0$. *Put*

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

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

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

Using Newton's binomium (Exercise 15 on page 14), which holds in any commutative ring, we have:

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

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

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

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

In case K is finite every injective map from K to itself (so for example F) is surjective as well. Hence F is bijective. The fact that F^{-1} is a field homomorphism as well, is easily verified. This shows Theorem VIII.1.1. ∎

VIII.1.2 Example. Let $K = \mathbb{F}_p(T)$ be the field of rational functions (= quotients of polynomials) with coefficients in $\mathbb{F}_p$. Put $f(T) = \frac{a_0+a_1T+\ldots+a_nT^n}{b_0+b_1T+\ldots+b_mT^m} \in \mathbb{F}_p(T)$, then

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

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

VIII.1.3 Definition. If L and L' are extensions of a field K then a *K-homomorphism* $L \to L'$ is a field homomorphism

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

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

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

VIII.1.4 Definition. Given an extension L of a field K we write $\mathrm{Aut}_K(L)$ for the group of all K-automorphisms of L. Here the group law is the composition of maps, and the unit element is id_L.

VIII.1.5 Theorem. *Let L be a finite extension of the field K and $\alpha \in L$ with minimal polynomial $f_K^\alpha \in K[X]$.*

(i) For every $\phi \in \mathrm{Aut}_K(L)$ we have that $\phi(\alpha)$ is a zero of f_K^α.

(ii) If m is the number of zeros of f_K^α in the field $K[\alpha] \subset L$, then $\#\mathrm{Aut}_K(K[\alpha]) = m$. Here $\#S$ the number of elements of a set S.

(iii) The number of K-automorphisms of $K[\alpha]$ is $\leq \deg(f_K^\alpha)$.

Proof. (i): Write $f_K^\alpha = X^n + a_{n-1}X^{n-1} \ldots + a_1 X + a_0$ with $a_i \in K$. Applying $\phi \in \mathrm{Aut}_K(L)$ to the equality $0 = f_K^\alpha(\alpha)$ yields:

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

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

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

is a bijection.

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

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

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

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

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

the evaluation homomorphism in β. Since β is a zero of $f_K^\alpha \in K[X]$ we have $f_K^\alpha \in \mathrm{Ker}(\mathrm{ev}_\beta)$. However f_K^α is irreducible in $K[X]$, hence $\mathrm{Ker}(\mathrm{ev}_\beta) = (f_K^\alpha)$. So ev_β induces an injective K-homomorphism $\overline{\mathrm{ev}_\beta} : K[X]/(f_K^\alpha) \to K[\alpha]$ with image $K[\beta] \subset K[\alpha]$. The first isomorphism theorem for rings II.3.7 now implies $K[\beta] \cong K[X]/(f_K^\alpha)$; the given isomorphism is K-linear, hence $K[\beta]$ is a linear subspace of $K[\alpha]$ with dimension $\dim_K(K[X]/(f_K^\alpha)) = \dim_K(K[\alpha])$. This dimension is finite and therefore $K[\beta] = K[\alpha]$; in other words $\overline{\mathrm{ev}_\beta}$ is an isomorphism. Analogous to the above, write $\overline{\mathrm{ev}_\alpha}$ for the isomorphism $K[X]/(f_K^\alpha) \xrightarrow{\cong} K[\alpha]$ induced by the evaluation homomorphism ev_α. We have isomorphisms:

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

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

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

hence $\phi(\alpha) = \beta$.
(iii) is immediate from (ii) and Theorem III.5.2. This finishes the proof. ∎

VIII.1.6 Example. Let $f \in K[X]$ be a monic irreducible polynomial of degree 2 and put $L = K[X]/(f)$. We write $\alpha := X + (f) \in L$ which is a zero of f in $L = K[\alpha]$. Writing $f = X^2 + aX + b$ one calculates (using long division if necessary) that in $L[X]$:

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

We now distinguish two cases:

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

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

In the remaining case (ii) f has two distinct zeros in L, hence $\#\mathrm{Aut}_K(L) = 2$ by Theorem VIII.1.5. All groups consisting of two elements are isomorphic, so

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

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

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

Another example is given by the finite fields $L = \mathbb{F}_p[X]/(X^2 - d)$ with d a non-square in $\mathbb{F}_p$ (see V.2.9, in particular we must have $p > 2$ here). Then

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

is the nontrivial field automorphism.

On the other hand Theorem VIII.1.1 yields an automorphism as well, namely the Frobenius automorphism F. We have

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

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

$$F = \phi.$$

As $\#\mathrm{Aut}_{\mathbb{F}_p}(L) = 2$ it suffices to show $F \neq id_L$. In case $F = \mathrm{id}_L$ the polynomial $X^p - X$ would have $\#L = p^2$ zeros in L. This contradicts Theorem III.5.2. Conclusion: $F \neq \mathrm{id}_L$ and therefore $F = \phi$.

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

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

This shows $d^{\frac{p-1}{2}} = -1$ in $\mathbb{F}_p \subset L$. We will say more about this in Corollary IX.4.4 below. ∎

VIII.2 Splitting fields

As we saw, for any irreducible $f \in K[X]$ there exists a finite extension $M \supseteq K$ in which f has a zero. Namely, $M = K[X]/(f)$ works; see Theorem V.2.7. In $M[X]$ one can write $f = (X - \alpha)g$. In order to obtain an extension of K in which f has (counted

with multiplicity) precisely $\deg(f)$ zeros, factor g into irreducible polynomials in $M[X]$. If an irreducible factor of degree ≥ 2 occurs, use it to construct a further finite extension (of M and therefore of K) in which this factor has a zero. After finitely many steps one obtains in this way a finite extension L of K such that f factors in $L[X]$ as a product

$$f = (X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n), \qquad \alpha_i \in L.$$

We elaborate on this in (the proof of) Theorem VIII.2.3.

VIII.2.1 Definition. Let K be a field and $f \in K[X]$ a monic polynomial. An extension $L \supseteq K$ is called a *splitting field* of f over K if $\alpha_1, \ldots, \alpha_n \in L$ exist with

(i) $f = \prod_{i=1}^{n}(X-\alpha_i)$ in $L[X]$,
(ii) $L = K(\alpha_1, \ldots, \alpha_n)$.

Roughly speaking: a splitting field of f over K is obtained by adjoining 'all' zeros of f to the field K (however, zeros from *where*?). Note that a splitting field of f is *finite* over K, as follows from (ii) in the definition together with Corollary VII.3.7.

VIII.2.2 Example. Take $f = X^3 - n$ with $n \in \mathbb{Q}$ and $n \neq k^3$ for all $k \in \mathbb{Q}$. We describe a splitting field M of f over $\mathbb{Q}$ and we determine $\mathrm{Aut}(M) = \mathrm{Aut}_{\mathbb{Q}}(M)$.

The polynomial f is irreducible in $\mathbb{Q}[X]$ since it has degree 3 and it has no zero in $\mathbb{Q}$ by the choice of n. Put

$$L := \mathbb{Q}[X]/(f), \qquad \alpha_1 := X + (f) \in L,$$

then L is a field and $\alpha_1 \in L$ is a zero of f. In particular $\alpha_1^3 = n$ and $[L:\mathbb{Q}] = 3$.

In $L[X]$ one obtains the factorization

$$f(X) = X^3 - n = (X-\alpha_1)(X^2 + \alpha_1 X + \alpha_1^2) =: (X-\alpha_1)g(X).$$

We now show that $g(X)$ is irreducible in $L[X]$. For this, it suffices to verify that g has no zero in L. Were $\beta \in L$ a zero of g then $0 = \beta^2 + \alpha_1\beta + \alpha_1^2 = \alpha_1^2\left((\frac{\beta}{\alpha_1})^2 + (\frac{\beta}{\alpha_1}) + 1\right)$ hence $\frac{\beta}{\alpha_1} \in L$ would be a zero of $X^2 + X + 1$. Since $X^2 + X + 1 \in \mathbb{Q}[X]$ is monic and irreducible, one concludes it is the minimal polynomial of $\frac{\beta}{\alpha_1}$ over $\mathbb{Q}$. This contradicts $[L:\mathbb{Q}] = 3$; compare Exercise 12 in Chapter VII. We conclude that $g(X)$ is irreducible in $L[X]$.

Now we adjoin a zero of g to L. Put

$$M := L[Y]/(g) \qquad \text{and} \qquad \alpha_2 := Y + (g) \in M.$$

(We use a variable Y instead of X to avoid confusion(!)) Then M is a degree 2 extension of L and therefore by Theorem VII.3.5 a degree 6 extension of $\mathbb{Q}$. By construction $\alpha_2 \in M$ is a zero of g and hence also of f. In $M[Z]$ we have the factorization $g(Z) = Z^2 + \alpha_1 Z + \alpha_1^2 = (Z-\alpha_2)(Z-\alpha_3)$ with $\alpha_3 = -\alpha_1 - \alpha_2$. Note that $\alpha_2 \neq \alpha_3$ since $\alpha_2 \in M$ is not a zero of the derivative of f (compare Theorem III.6.7). Hence f has three distinct zeros $\alpha_1, \alpha_2, \alpha_3$ in M and $f = (X-\alpha_1)(X-\alpha_2)(X-\alpha_3)$ in $M[X]$.

Since $M = L(\alpha_2) = L(\alpha_2, \alpha_3)$ and $L = \mathbb{Q}(\alpha_1)$ we obtain:

$$M = \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3),$$

which shows that M is a splitting field of f over $\mathbb{Q}$.

In fact even $M = \mathbb{Q}[\alpha_1, \alpha_2]$ and any $m \in M$ can be written as

$$m = a_0 + a_1\alpha_1 + a_2\alpha_1^2 + (b_0 + b_1\alpha_1 + b_2\alpha_1^2)\alpha_2,$$

with $a_i, b_i \in \mathbb{Q}$ (check for yourself!).

We will now show that

$$\mathrm{Aut}_{\mathbb{Q}}(M) \cong S_3,$$

with S_3 the group of permutations of $\{1,2,3\}$. Note that $\mathrm{Aut}(M) = \mathrm{Aut}_{\mathbb{Q}}(M)$ since $\mathbb{Q}$ is the prime field. Any $\phi \in \mathrm{Aut}_{\mathbb{Q}}(M)$ permutes the three zeros of f in M. This yields a group homomorphism

$$\pi : \mathrm{Aut}_{\mathbb{Q}}(M) \longrightarrow S_3, \qquad \pi : \phi \mapsto \sigma,$$

with $\sigma \in S_3$ defined by $\phi(\alpha_i) = \alpha_{\sigma(i)}$. We claim that π is injective and surjective, and this of course determines $\mathrm{Aut}_{\mathbb{Q}}(M)$.

Suppose $\phi \in \mathrm{Ker}(\pi)$, then $\phi(\alpha_i) = \alpha_i$ for all i. Applying ϕ to $m \in M$ then the above description of m implies $\phi(m) = m$, hence $\phi = \mathrm{id}_M$. We conclude that π is injective.

To show surjectivity of π we construct some field automorphisms of M. Using $\mathbb{Q} \subset L \subset M$ we have that $\mathrm{Aut}_L(M) \subset \mathrm{Aut}_{\mathbb{Q}}(M)$ is a subgroup. Now $M = L[\alpha_2]$ and the minimal polynomial g of α_2 over L has two distinct zeros in M, hence Theorem VIII.1.5 shows that $\mathrm{Aut}_L(M)$ contains exactly one element $\phi_1 \neq \mathrm{id}_M$. As ϕ_1 is the identity on L, we have $\phi_1(\alpha_1) = \alpha_1$. Moreover $\phi_1 \neq \mathrm{id}_M$ implies that ϕ_1 interchanges α_2 and α_3. Conclusion:

$$\phi_1(\alpha_1) = \alpha_1, \quad \phi_1(\alpha_2) = \alpha_3, \quad \phi_1(\alpha_3) = \alpha_2 \quad \Longrightarrow \quad \pi(\phi_1) = (23) \in S_3.$$

Now put $L_2 := \mathbb{Q}(\alpha_2) \subset M$. Since $\alpha_2 \in L_2$ is a zero of the irreducible $f \in \mathbb{Q}[X]$ one concludes $L_2 \cong \mathbb{Q}[X]/(f)$, and the reasoning above can be repeated with L replaced by L_2. This results in $\phi_2 \in \mathrm{Aut}_{L_2}(M) \subset \mathrm{Aut}_{\mathbb{Q}}(M)$ with

$$\phi_2(\alpha_1) = \alpha_3, \quad \phi_2(\alpha_2) = \alpha_2, \quad \phi_2(\alpha_3) = \alpha_1 \quad \Longrightarrow \quad \pi(\phi_2) = (13) \in S_3.$$

Since (13) and (23) generate S_3 this shows that π is surjective, completing the proof of the assertion that $\mathrm{Aut}(M) \cong S_3$. ∎

VIII.2.3 Theorem. *Let K be a field and $f \in K[X]$ a monic polynomial.*
There exists a splitting field of f over K.

Proof. We use induction w.r.t. $n = \deg(f)$. If $n = 1$ then K itself is a splitting field of f over K. Now take $n > 1$. We distinguish two cases: f is irreducible or not.

First suppose f can be factored: $f = g \cdot h$ with $g, h \in K[X]$ monic of degree $< n$. The induction hypothesis yields that a splitting field $E = K(\beta_1, \beta_2, ..., \beta_m)$ of g over K exists, and $g = \prod_{i=1}^{m}(X - \beta_i)$ in $E[X]$. Moreover the induction hypothesis applied to the field E and the polynomial $h \in E[X]$ yields a splitting field $L = E(\gamma_1, \gamma_2, ...\gamma_k)$ of h over E, and $h = \prod_{i=1}^{k}(X - \gamma_i)$ in $L[X]$. Then L is a splitting field of f over K, since $f = \prod_{i=1}^{m}(X - \beta_i) \cdot \prod_{i=1}^{k}(X - \gamma_i)$ in $L[X]$ and $L = E(\gamma_1, \gamma_2, ..., \gamma_k) = K(\beta_1, ..., \beta_m, \gamma_1, ..., \gamma_k)$.

Next suppose that f is irreducible in $K[X]$. Then by V.2.7 an extension $K(\alpha)$ exists with $f(\alpha) = 0$. Using III.5.1 $h \in K(\alpha)[X]$ exists with $f = (X - \alpha)h$. Here h is monic and of degree $n - 1$. Applying the induction hypothesis to $K(\alpha)$ and h one obtains a splitting field $L = K(\alpha)(\alpha_1, ..., \alpha_{n-1})$ of h over $K(\alpha)$, and $h = \prod_{i=1}^{n-1}(X - \alpha_i)$ in $L[X]$. With $\alpha_n = \alpha$ we now have $L = K(\alpha_1, \alpha_2, ..., \alpha_n)$ and $f = \prod_{i=1}^{n}(X - \alpha_i)$ in $L[X]$, so L is a splitting field of f over K. This proves VIII.2.3. ∎

We will now show the *unicity* of the splitting field. First a quite general result is discussed.

VIII.2.4 Theorem. *Let $\phi : K_0 \to K_1$ be an isomorphism from a field K_0 to a field K_1 and*

$$\Phi : K_0[X] \longrightarrow K_1[X], \qquad \sum_i a_i X^i \mapsto \sum_i \phi(a_i) X^i$$

the isomorphism of polynomial rings induced by ϕ. Take $f_0 \in K_0[X]$ monic and let L_0 be a splitting field of f_0 over K_0 and L_1 a splitting field of $f_1 := \Phi(f_0) \in K_1[X]$ over K_1.

Then there is an isomorphism $\psi : L_0 \to L_1$ extending the map ϕ:

$$\psi : L_0 \xrightarrow{\sim} L_1, \qquad \psi_{|K_0} = \phi : \quad \begin{array}{ccc} K_0 & \hookrightarrow & L_0 \\ \phi \downarrow & & \downarrow \psi \\ K_1 & \hookrightarrow & L_1. \end{array}$$

Proof. From Definition VIII.2.1(ii) and Corollary VII.3.7 it follows that $[L_0 : K_0]$ is finite. We use induction w.r.t. this degree.

If $[L_0 : K_0] = 1$ then $L_0 = K_0$, hence $f_0 = \prod_{i=1}^n (X - \beta_i)$ with $\beta_1, \ldots, \beta_n \in K_0$. Then $f_1 = \Phi(f_0) = \prod_{i=1}^n \Phi(X - \beta_i) = \prod_{i=1}^n (X - \phi(\beta_i)) \in K_1[X]$. All zeros of f_1 in L_1 are therefore in K_1, and since L_1 is obtained by adjoining these zeros to K_1 we have $L_1 = K_1$. Hence we may take $\psi = \phi$.

Now assume $[L_0 : K_0] > 1$. We construct simple extensions $K_0(\alpha_0)$ and $K_1(\alpha_1)$ and an isomorphism $\chi : K_0(\alpha_0) \to K_1(\alpha_1)$ with $\chi_{|K_0} = \phi$.

As $[L_0 : K_0] > 1$ we can take $\alpha_0 \in L_0$ with $f_0(\alpha_0) = 0$ and $\alpha_0 \notin K_0$. Let $h_0 \in K_0[X]$ be the minimal polynomial of α_0 over K_0 and $h_1 := \Phi(h_0) \in K_1[X]$. Then h_0 is a divisor of f_0 in $K_0[X]$, so

$$f_0 = h_0 q_0 \in K_0[X] \quad \text{and therefore} \quad f_1 = h_1 \Phi(q_0) \in K_1[X].$$

However, f_1 is in $L_1[X]$ a product of linear factors, hence the same holds for $h_1 \mid f_1$. We conclude that h_1 has a zero $\alpha_1 \in L_1$. Claim: $K_0(\alpha_0) \cong K_1(\alpha_1)$.

Indeed, by VII.2.5 one obtains the isomorphism

$$K_0[X]/(h_0) \xrightarrow{\cong} K_0(\alpha_0), \qquad X + (h_0) \mapsto \alpha_0. \tag{1}$$

Since $h_0 \in K_0[X]$ is irreducible and Φ is an isomorphism, $h_1 = \Phi(h_0)$ is irreducible in $K_1[X]$. As $h_1(\alpha_1) = 0$ one concludes that h_1 is the minimal polynomial of α_1 over K_1. Again by VII.2.5 one finds the isomorphism

$$K_1[X]/(h_1) \xrightarrow{\cong} K_1(\alpha_1), \qquad X + (h_1) \mapsto \alpha_1. \tag{2}$$

Finally, the isomorphism $\Phi : K_0[X] \xrightarrow{\simeq} K_1[X]$ maps the ideal generated by h_0 to the ideal generated by $h_1 = \Phi(h_0)$, hence Φ induces an isomorphism

$$\overline{\Phi} : K_0[X]/(h_0) \xrightarrow{\cong} K_1[X]/(h_1), \qquad X + (h_0) \mapsto X + (h_1). \tag{3}$$

The restriction of $\overline{\Phi}$ to K_0 equals ϕ. Combining the isomorphisms (1), (2), and (3) one obtains

$$\chi : K_0(\alpha_0) \xrightarrow{\cong} K_1(\alpha_1), \qquad \alpha_0 \mapsto \alpha_1, \quad \chi_{|K_0} = \phi.$$

To finish the proof, we will show that we can apply the induction hypothesis to $\chi : K_0(\alpha_0) \to K_1(\alpha_1)$ and their respective extensions L_0, L_1. We pick α_0 not in K_0 hence $[K_0(\alpha_0) : K_0] > 1$ and therefore

$$[L_0 : K_0(\alpha_0)] = \frac{[L_0 : K_0]}{[K_0(\alpha_0) : K_0]} < [L_0 : K_0].$$

Moreover L_0 is a splitting field of f_0 over $K_0(\alpha_0)$, and analogously L_1 is a splitting field of f_1 over $K_1(\alpha_1)$: indeed, adjoining all zeros of f_0 in L_0 to K_0 one obtains L_0,

hence this certainly holds if we adjoin them to $K_0(\alpha_0)$. The same reasoning holds for $K_1(\alpha_1)$ and f_1 and L_1.

The induction hypothesis therefore yields a field isomorphism $\psi : L_0 \to L_1$ with $\psi_{|K_0(\alpha_0)} = \chi$. In other words, we obtain the rightmost part of the diagram

$$\begin{array}{ccccc} K_0 & \hookrightarrow & K_0(\alpha_0) & \hookrightarrow & L_0 \\ \phi\downarrow & & \chi\downarrow & & \downarrow\psi \\ K_1 & \hookrightarrow & K_1(\alpha_1) & \hookrightarrow & L_1. \end{array}$$

In particular $\psi_{|K_0} = \phi$, finishing the proof of VIII.2.4. ∎

A consequence of the above is the main result on splitting fields, which we now state and prove.

VIII.2.5 Theorem. *Let K be a field and $f \in K[X]$ monic.*

There exists a splitting field of f over K, and this splitting field is unique up to K-isomorphisms.

Proof. The existence was shown in VIII.2.3. Suppose L and L' are splitting fields of f over K; we must show that a K-isomorphism $\psi : L \to L'$ exists. Applying VIII.2.4 to $K_0 = K_1 = K$, $f_0 = f_1 = f$, $\phi = id_K$, $L_0 = L, L_1 = L'$ this is immediate. Hence VIII.2.5 is proven. ∎

VIII.2.6 Notation. The (unique by Theorem VIII.2.5) splitting field of f over K is denoted by Ω_K^f.

VIII.2.7 Remark. The K-isomorphism $\psi : L \to L'$ between two splitting fields of f over K is in general not unique. If σ is a K-automorphism of L then $\psi' = \psi \circ \sigma : L \to L'$ is also a K-isomorphism as desired. It is not hard to show that starting from a fixed ψ *all* possible K-isomorphisms $\psi' : L \to L'$ are obtained as above by composing with elements of $\mathrm{Aut}_K(L)$.

VIII.2.8 Example. We return to Example VIII.2.2. Here $f = X^3 - n$ and in $\mathbb{C}[X]$ one has

$$f(X) = (X - \beta_1)(X - \beta_2)(X - \beta_3) \qquad \text{m}$$

$$\beta_1 = \sqrt[3]{n}, \quad \beta_2 = \sqrt[3]{n}(-\frac{1}{2} + i\frac{\sqrt{3}}{2}), \quad \beta_3 = \sqrt[3]{n}(-\frac{1}{2} - i\frac{\sqrt{3}}{2}).$$

Hence $\mathbb{Q}(\beta_1, \beta_2, \beta_3)$ $(\subset \mathbb{C})$ is another splitting field of f over $\mathbb{Q}$. By Theorem VIII.2.5 it is isomorphic to the one studied in Example VIII.2.2: $M = \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3)$. In particular

$$[\mathbb{Q}(\beta_1, \beta_2, \beta_3) : \mathbb{Q}] = 6 \qquad \text{and} \quad \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\beta_1, \beta_2, \beta_3)) \cong S_3.$$

—∎

VIII.3 Exercises

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

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

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

2. Describe a splitting field of $X^2 - 101$ over $\mathbb{Q}$.
3. Let L be a splitting field of f over K and $f = \prod_{i=1}^{n}(X - \alpha_i)$ in $L[X]$. Prove that $L = K(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$ (so α_n is omitted!).
4. Suppose $f \in K[X]$ is monic of degree n. Prove: $[\Omega_K^f : K]$ divides $n!$ (Hint: use the construction in the proof of VIII.2.3.)
5. Prove that $L = \mathbb{Q}(\sqrt[4]{2}, i)$ is a splitting field of $X^4 - 2$ over $\mathbb{Q}$. Determine $[L : \mathbb{Q}]$ and $\sharp\mathrm{Aut}(L)$.
   Prove that $\mathrm{Aut}_{\mathbb{Q}(i)}(L) \cong \mathbb{Z}/4\mathbb{Z}$.
6. Let $\zeta \in \mathbb{C}$ be a zero of $f = X^4 + X^3 + X^2 + X + 1$. Show that $\zeta^5 = 1$ and that $\zeta^2, \zeta^3, \zeta^4$ are zeros of f as well. Prove that $\mathbb{Q}(\zeta)$ is a splitting field of f over $\mathbb{Q}$. Determine $\mathrm{Aut}(\mathbb{Q}(\zeta))$.
7. Prove that $\Omega_{\mathbb{Q}}^{X^2-2} \ncong \Omega_{\mathbb{Q}}^{X^2-3}$ and that $\Omega_K^{X^2-\bar{2}} \simeq \Omega_K^{X^2-\bar{3}}$ for $K = \mathbb{F}_5$.
8. Prove that $\mathbb{Q}(i, \sqrt{2})$ is a splitting field of $f_{\mathbb{Q}}^{i+\sqrt{2}}$ over $\mathbb{Q}$.
   Show that $\mathrm{Aut}(\mathbb{Q}(i, \sqrt{2})) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.
9. Let L be a splitting field of f over K and put $n = \deg(f)$.
   (a) Prove: any K-automorphism of L permutes the zeros of f in L,
   (b) Prove: the group $\mathrm{Aut}_K(L)$ is isomorphic to a subgroup of S_n;
   (c) Show that $\#\mathrm{Aut}_K(L)$ is a divisor of $n!$.
10. Prove: $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3})) \simeq S_3$.
11. Take $f = X^3 + X^2 - 2X - 1 \in \mathbb{Q}[X]$ and let $\alpha \in \Omega_{\mathbb{Q}}^f$ be a zero of f.
    Compute $f_{\mathbb{Q}}^{\alpha^2-2}$ and show that $\mathbb{Q}(\alpha) = \Omega_{\mathbb{Q}}^f$. Deduce that $\mathrm{Aut}(\Omega_{\mathbb{Q}}^f) \cong \mathbb{Z}/3\mathbb{Z}$.

IX FINITE FIELDS

A field K is called *finite* if the number of elements of K is finite. We already saw some examples of finite fields: $\mathbb{Z}/p\mathbb{Z}$ for p prime, see Theorem I.2.11; we saw a field with 4 elements in Example III.3.6, and fields with p^2 elements, for any prime $p > 2$, were constructed in Example V.2.9.

IX.1 Classification of finite fields.

The next result classifies the finite fields.

IX.1.1 Theorem.
(i) If K is a finite field then $\#K = p^n$ for a prime p and an integer $n \geq 1$.
(ii) For every prime p and every integer $n \geq 1$ there exists a field consisting of p^n elements, namely the splitting field of $X^{p^n} - X$ over $\mathbb{F}_p$. This field is unique up to isomorphisms.

Proof. (i): Let K be a finite field. Using Theorem VII.1.3 one concludes that the prime field $K_0 \subseteq K$ must be $\mathbb{F}_p$ for a prime p. As K is finite, K is certainly finite dimensional as a vectorspace over $\mathbb{F}_p$. Put $n = [K : \mathbb{F}_p]$. Choosing a basis $e_1, e_2, \ldots, e_n$ for K over $\mathbb{F}_p$, every $x \in K$ can be written uniquely as

$$x = a_1 e_1 + a_2 e_2 + \ldots + a_n e_n \qquad \text{with} \quad a_i \in \mathbb{F}_p,\ 1 \leq i \leq n.$$

For each of the a_i there are p possibilities, hence in total K contains $p \cdot p \cdot \ldots \cdot p = p^n$ elements, proving (i).

(ii): Let p be prime and $n \in \mathbb{Z}_{>0}$ and $q = p^n$. Let K be the splitting field of $X^q - X$ over $\mathbb{F}_p$. We will show that $\#K = q$, as follows. Since K is the splitting field of $X^q - X$ over $\mathbb{F}_p$, we have $\alpha_1, \alpha_2, \ldots, \alpha_q \in K$ with $X^q - X = \prod_{i=1}^{q}(X - \alpha_i)$ in $K[X]$. We will study the set

$$A = \{\alpha_1, \alpha_2, \ldots, \alpha_q\} \qquad (\subset K).$$

(a.) $\#A = q$. Indeed, were $\#A < q$ then $\alpha_i = \alpha_j$ for some $i \neq j$. This means that α_i is a multiple zero of $f = X^q - X$. Theorem III.6.7 now implies that α_i is also a zero of the derivative $f' = q \cdot X^{q-1} - 1 = -1$ (note that $q = 0$ in $\mathbb{F}_p$). This is absurd since the constant polynomial -1 has no zeros. We conclude $\#A = q$.
(b.) A is a *subfield* of K. Indeed, by definition the set A consists of all zeros of $X^q - X$ in K. Hence $\alpha \in K$ satisfies: $\alpha \in A \iff \alpha^q = \alpha$. Hence clearly $1 \in A$. Furthermore:

$$\alpha, \beta \in A,\ \beta \neq 0 \Longrightarrow (\alpha\beta^{-1})^q = \alpha^q (\beta^q)^{-1} = \alpha\beta^{-1} \Longrightarrow \alpha\beta^{-1} \in A.$$

Using Theorem VIII.1.1 one obtains:

$$\alpha,\beta\in A\Longrightarrow(\alpha+\beta)^q=\alpha^q+\beta^q=\alpha+\beta\Longrightarrow\ \alpha+\beta\in A.$$

This shows that A is closed under addition and division. Since also $-1\in A$ it follows that A is a subfield of K.
(c.) $A=K$ $(:=\Omega_{\mathbb{F}_p}^{X^q-X})$. Indeed, since A is a subfield of K it necessarily contains the prime field $\mathbb{F}_p$ of K (by the definition of prime field). All $\alpha_i\in A$ hence

$$\mathbb{F}_p(\alpha_1,\alpha_2,\ldots,\alpha_q)\subset A.$$

The definition of a splitting field VIII.2.1 then shows $K=\mathbb{F}_p(\alpha_1,\alpha_2,\ldots,\alpha_q)$. We conclude that $K\subset A$ and therefore $K=A$.

This shows that a field consisting of q elements exists, namely the splitting field of X^q-X over $\mathbb{F}_p$. It remains to prove *unicity*.

Suppose L is another field consisting of q elements. We must show $L\cong K$. Observe that $\mathrm{char}(L)=p$ since otherwise by (i) above $\#L$ would be a power of a different prime, contradicting the unique prime factorization of q. Let $\alpha\in L^\times$. Since $L^\times$ is a group consisting of $q-1$ elements, the order of α in this group divides $q-1$. Therefore $\alpha^{q-1}=1$. This implies $\alpha^q=\alpha$ hence α is a zero of the polynomial X^q-X. Clearly also $\alpha=0$ has this property, so all q elements of L are zeros of X^q-X. This polynomial has degree q and hence $X^q-X=\prod_{\alpha\in L}(X-\alpha)$. Using Definition VIII.2.1 one concludes that L is a splitting field of X^q-X over the prime field $\mathbb{F}_p\subseteq L$. The unicity of splitting fields (see Theorem VIII.2.5) then implies $L\cong K$. This finishes the proof of Theorem IX.1.1. ∎

IX.1.2 Definition. The (unique by Theorem IX.1.1) field consisting of $q=p^n$ elements is denoted by $\mathbb{F}_q$.

So Theorem IX.1.1 says in particular that $\mathbb{F}_q$ is the splitting field of X^q-X over $\mathbb{F}_p$, i.e.,

$$\mathbb{F}_q\cong\Omega_{\mathbb{F}_p}^{X^q-X}.$$

IX.1.3 Remark. Instead of $\mathbb{F}_q$ the literature also uses the notation GF(q), for 'Galois field', named after Evariste Galois (French mathematician, 1811 - 1832) who was the first to study finite fields in general.

IX.1.4 Remark. If q is *prime* then $\mathbb{F}_q\cong\mathbb{Z}/q\mathbb{Z}$. However if q is *not* prime then $\mathbb{Z}/q\mathbb{Z}$ is not a field (Theorem I.2.11, note: for $q=p^n$ and $n>1$ one finds $\overline{p}\neq\overline{0}\neq\overline{p^{n-1}}$ and $\overline{p}\cdot\overline{p^{n-1}}=\overline{p^n}=\overline{0}$ in $\mathbb{Z}/q\mathbb{Z}$). So this ring contains zero divisors. Also, given any $a\in\mathbb{F}_q$ we have $a+a+\ldots+a$ (p terms) is equal to $0\in\mathbb{F}_q$, since $\mathbb{F}_q$ is a vector space over $\mathbb{F}_p$. On the other hand in $\mathbb{Z}/q\mathbb{Z}$ we have that $\overline{1}+\overline{1}+\ldots+\overline{1}$ (p terms) is *not* zero whenever $q=p^n>p$.

So clearly $\mathbb{F}_q\ncong\mathbb{Z}/q\mathbb{Z}$ whenever $q=p^n>p$.

IX.1.5 Example. The polynomials $f:=X^3+X^2+1$ and $g:=X^3+X+1$ are irreducible in $\mathbb{F}_2[X]$ since they have degree 3 and they have no zero in $\mathbb{F}_2$. The fields

$$K:=\mathbb{F}_2[X]/(f)\qquad\text{and}\qquad L:=\mathbb{F}_2[X]/(g)$$

both consist of 8 elements hence by Theorem IX.1.1 $K\cong L$. We now construct an explicit isomorphism.

Note that $\alpha:=X+(f)$ is a zero of f in K. Since K and L are isomorphic, f must have a zero in L as well, and this is what we will find first. Put

$$\beta:=X+(g)\in L,\qquad\text{then}\qquad\beta^3=\beta+1.$$

Then

$$\begin{aligned} f(\beta+1) &= (\beta+1)^3+(\beta+1)^2+1 \\ &= (\beta^3+\beta^2+\beta+1)+(\beta^2+1)+1 \\ &= \beta^2+(\beta^2+1)+1 \\ &= 0, \end{aligned}$$

(note that the coefficients are in $\mathbb{F}_2$), so we have our zero. Since it turns out to be rather elaborate to describe a field isomorphism from K to L directly, one uses the evaluation homomorphism

$$\mathrm{ev}_{\beta+1} : \mathbb{F}_2[X] \longrightarrow L, \qquad X \mapsto \beta+1.$$

It is not hard to verify that $\mathrm{ev}_{\beta+1}$ is surjective. The kernel of $\mathrm{ev}_{\beta+1}$ is generated by f (it is in the kernel and it is irreducible). The first isomorphism theorem II.3.7 now yields

$$K = \mathbb{F}_2[X]/(f) \cong \mathrm{ev}_{\beta+1}(\mathbb{F}_2[X]) = L,$$

with explicit isomorphism given by

$$a_0+a_1\alpha+a_2\alpha^2 \mapsto a_0+a_1(\beta+1)+a_2(\beta+1)^2 = (a_0+a_1+a_2)+a_1\beta+a_2\beta^2.$$

—■

IX.2 The structure of finite fields

Since the finite field $\mathbb{F}_{p^n}$ is an n-dimensional vectorspace over $\mathbb{F}_p$ it follows that the additive group $(\mathbb{F}_{p^n},+,0)$ is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$ (a product of n copies of $\mathbb{Z}/p\mathbb{Z}$). The next result shows that $\mathbb{F}_{p^n}$ is a simple extension of $\mathbb{F}_p$ (see Definition VII.2.4), which allows one to treat this field analogously to Example VII.2.10.

IX.2.1 Theorem. *Let $\mathbb{F}_q$ be a finite field and $q = p^n$ with p prime. Then $\alpha \in \mathbb{F}_q$ exists with*

$$\mathbb{F}_q = \mathbb{F}_p[\alpha].$$

In particular $\mathbb{F}_q \cong \mathbb{F}_p[X]/(f^\alpha_{\mathbb{F}_p})$ and $\deg(f^\alpha_{\mathbb{F}_p}) = n$.

Proof. For every $\beta \in \mathbb{F}_q$ the field $\mathbb{F}_p(\beta)$ is finite hence $\mathbb{F}_p(\beta) \cong \mathbb{F}_{p^k}$ for some $k \le n$. Hence β is a zero of $X^{p^k}-X$, a polynomial having at most (in fact, exactly) p^k pairwise distinct zeros in $\mathbb{F}_{p^n}$. Considering all possible k, we see that the number of elements $\beta \in \mathbb{F}_{p^n}$ such that $\mathbb{F}_p(\beta) \neq \mathbb{F}_{p^n}$ is at most $p+p^2+\ldots+p^{n-1} < p^n$. Hence $\alpha \in \mathbb{F}_{p^n}$ exists with $\mathbb{F}_p(\alpha) = \mathbb{F}_{p^n}$.

The remaining statements are immediate from Theorem VII.2.5. ■

In Theorem IX.2.7 below we will see a complete description of all subfields of $\mathbb{F}_q$.

IX.2.2 Corollary. *For any p prime and any integer $n > 0$ an irreducible polynomial of degree n in $\mathbb{F}_p[X]$ exists.*

Proof. By Theorem IX.2.1 $\alpha \in \mathbb{F}_{p^n}$ exists with $\mathbb{F}_{p^n} = \mathbb{F}_p[\alpha]$. The minimal polynomial $f^\alpha_{\mathbb{F}_p}$ of α over $\mathbb{F}_p$ is then irreducible of degree $[\mathbb{F}_{p^n} : \mathbb{F}_p] = n$. This proves the corollary. ■

IX.2.3 Example. To be able to compute in $\mathbb{F}_{5^3}$ we first look for an irreducible polynomial of degree 3 in $\mathbb{F}_5[X]$. The polynomials of the form $X^3 - a$ with $a \in \mathbb{F}_5$ all have a zero in $\mathbb{F}_5$ (verify!) so these are reducible. $f = X^3 + X - 1 \in \mathbb{F}_5[X]$, turns out to be irreducible (again, verify!). Hence $L = \mathbb{F}_5[X]/(f)$ is a field and $[L : \mathbb{F}_5] = 3$ so $\#L = 5^3$: every element of L can be given uniquely as

$$x = a_0 + a_1\alpha + a_2\alpha^2, \qquad \text{with} \quad \alpha := X + (f) \in L$$

and $a_i \in \mathbb{F}_5$. By Theorem IX.1.1 and Definition IX.1.2 we have $L \cong \mathbb{F}_{5^3}$. As α is a zero of f one finds $\alpha^3 = -\alpha + 1$. Hence for example

$$\begin{aligned}(3\alpha + 1)(4\alpha^2 + 2) &= 2\alpha^3 + 4\alpha^2 + \alpha + 2\\ &= 2(-\alpha + 1) + 4\alpha^2 + \alpha + 2\\ &= 4\alpha^2 + 4\alpha + 4,\\ &= -(\alpha^2 + \alpha + 1).\end{aligned}$$

where we note that the coefficients are taken in $\mathbb{F}_5$. ∎

Let $f \in \mathbb{F}_q[X]$ be irreducible. Then f has a zero α in the extension $L = \mathbb{F}_q[X]/(f)$ where $\#L = q^m$ with $m = \deg(f)$. Hence writing $q = p^n$ we have $L \cong \mathbb{F}_{p^{nm}}$. In $L[X]$ the polynomial f factors as $(X - \alpha)g$ for some $g \in L[X]$. Contrary to what we saw in Example VIII.2.2, it turns out that g (and f) split completely in $L[X]$ (Theorem IX.2.4 below), so L is the splitting field of f over $\mathbb{F}_q$. This result will allow us (Theorem IX.2.6 below) to determine $\mathrm{Aut}(\mathbb{F}_q)$. To this end, we use the Frobenius homomorphism (see VIII.1.1)

$$F : L \longrightarrow L \qquad x \mapsto x^p.$$

Since L is finite, F is a field automorphism, i.e., $F \in \mathrm{Aut}(L)$. For $k \in \mathbb{N}$ composing k Frobenius homomorphisms yields

$$F^k : L \longrightarrow L, \qquad x \mapsto x^{p^k}, \qquad F^k \in \mathrm{Aut}(L).$$

IX.2.4 Theorem. *Let $q = p^n$ be a power of a prime p and $f \in \mathbb{F}_q[X]$ monic and irreducible of degree m. Put $L := \mathbb{F}_q[X]/(f)$ and take $\alpha \in L$ with $f(\alpha) = 0$.*

Then

$$f = (X - \alpha)(X - \alpha^q)\cdots(X - \alpha^{q^{m-1}}) \quad \in L[X].$$

Moreover the m zeros of f are pairwise distinct.

Proof. For $a \in \mathbb{F}_q$ we have $F^n(a) = a^{p^n} = a^q$ and a is a zero of $X^q - X$. Therefore $F^n(a) = a$ hence $F^n \in \mathrm{Aut}_{\mathbb{F}_q}(L)$ since F^n is the identity map on $\mathbb{F}_q$.

For any $k \in \mathbb{N}$ then also $(F^n)^k \in \mathrm{Aut}_{\mathbb{F}_q}(L)$. Hence by Theorem VIII.1.5 also $(F^n)^k(\alpha) = \alpha^{q^k} \in L$ is a zero of f. We claim that in this way m distinct zeros of f are obtained. Namely, if $\alpha^{q^k} = \alpha^{q^l}$ with $0 \le k < l \le m - 1$ then (using $\mathrm{char}(L) = p$) one finds

$$0 = \alpha^{q^l} - \alpha^{q^k} = (\alpha^{q^{l-k}} - \alpha)^{q^k}.$$

This shows α would be a zero of $X^{q^a} - X$ where $a = l - k < m$. By Theorem IX.1.1 this means $\alpha \in \mathbb{F}_{q^a}$, contradicting $\mathbb{F}_q(\alpha) = \mathbb{F}_{q^m}$. So indeed the m zeros $\alpha, \alpha^q, \ldots, \alpha^{q^{m-1}} \in L$ of f are pairwise distinct and f factors in $L[X]$ in the given way. This completes the proof. ∎

IX.2.5 Example. We saw in Example IX.1.5 that the polynomial $f := X^3 + X^2 + 1$ has the zero $\beta + 1 \in L := \mathbb{F}_2[\beta] \cong \mathbb{F}_8$ with $\beta^3 = \beta + 1$. By Theorem IX.2.4 also

$$(\beta + 1)^2 = \beta^2 + 1 \qquad \text{and} \qquad (\beta + 1)^4 = \beta^4 + 1 = \beta^2 + \beta + 1$$

are zeros of f and hence

$$X^3+X^2+1=(X-(\beta+1))(X-(\beta^2+1))(X-(\beta^2+\beta+1)).$$

IX.2.6 Theorem. *Let $q=p^n$ with p prime. Then*

$$\mathrm{Aut}(\mathbb{F}_q)=\langle F\rangle\cong\mathbb{Z}/n\mathbb{Z},$$

so $\mathrm{Aut}(\mathbb{F}_q)$ *is a cyclic group consisting of n elements, with generator the Frobenius homomorphism.*

Proof. By Theorem IX.2.1 $\alpha\in\mathbb{F}_q$ exists with $\mathbb{F}_q=\mathbb{F}_p(\alpha)$ and by Theorem IX.2.4 the minimal polynomial of α has exactly n pairwise distinct zeros in $\mathbb{F}_q$. Hence Theorem VIII.1.5 implies that $\#\mathrm{Aut}(\mathbb{F}_q)=\#\mathrm{Aut}_{\mathbb{F}_p}(\mathbb{F}_q)=n$ (note $\mathbb{F}_p$ is the prime field of $\mathbb{F}_q$).

Next, observe that the n field automorphisms $\mathrm{id}_{\mathbb{F}_q},F,F^2,\ldots,F^{n-1}$ are all distinct: we saw in the proof of IX.2.4 that $F^k(\alpha)=\alpha^{p^k}\neq\alpha^{p^l}=F^l(\alpha)$ for $0\leq k<l<n-1$. Conclusion: every element of $\mathrm{Aut}(\mathbb{F}_q)$ is a power of the Frobenius automorphism. This shows the result. ∎

IX.2.7 Theorem. *Let $q=p^k$ and $r=p^m$. Then $\mathbb{F}_q$ is isomorphic to a subfield of $\mathbb{F}_r$ if and only if r is a power of q (in other words, precisely when $k|m$).*

In case $\mathbb{F}_q$ is isomorphic to a subfield of $\mathbb{F}_r$ then there is only one such subfield of $\mathbb{F}_r$, namely the set of zeros in $\mathbb{F}_r$ of X^q-X.

Proof. If $\mathbb{F}_q$ is a subfield of $\mathbb{F}_r$ then $\mathbb{F}_r$ is a finite dimensional vectorspace over $\mathbb{F}_q$ with $\dim_{\mathbb{F}_q}\mathbb{F}_r=[\mathbb{F}_r:\mathbb{F}_q]$. Therefore

$$p^m=\#\mathbb{F}_r=(\#\mathbb{F}_q)^{[\mathbb{F}_r:\mathbb{F}_q]}=p^{k[\mathbb{F}_r:\mathbb{F}_q]},\qquad\text{hence}\quad k|m.$$

Now assume $m=kn$ for some $n\in\mathbb{Z}_{>0}$, so $r=q^n$. We will show that the polynomial X^q-X splits completely in $\mathbb{F}_r[X]$. Since every element of $\mathbb{F}_r$ is a zero of X^r-X, it suffices to show that

$$X^q-X\,\Big|\,X^r-X\qquad(\text{in }\mathbb{F}_p[X]).$$

Both polynomials have a factor X and because $r=q^n$ it suffices to show

$$X^{q-1}-1\,\Big|\,X^{q^n-1}-1.$$

Note that in $\mathbb{Z}$ one has

$$q^n-1=(q-1)(q^{n-1}+q^{n-2}+\ldots+q+1)=(q-1)b$$

with $b\in\mathbb{Z}$. Hence the special case $a=q-1$, $b=q^{n-1}+q^{n-2}+\ldots+q+1$ of the equality

$$X^{ab}-1=(X^a-1)\left(X^{a(b-1)}+X^{a(b-2)}+\ldots+X^a+1\right)$$

shows the divisibility of the polynomials.

This proves the first part of IX.2.7. If $K\subseteq\mathbb{F}_r$ is a subfield consisting of q elements, then $x^q-x=0$ for all $x\in K$ hence K is the set of zeros in $\mathbb{F}_r$ of X^q-X. In particular at most one subfield consisting of q elements exists in $\mathbb{F}_r$. This finishes the proof. ∎

IX.2.8 Remark. We write, following Theorem IX.2.7:

$$\mathbb{F}_{p^k}\subset\mathbb{F}_{p^m}\iff\quad k|m.$$

Note, as an example, that $\mathbb{F}_4$ is not a subfield of $\mathbb{F}_8$. Both $\mathbb{F}_4$ and $\mathbb{F}_8$ are subfields of $\mathbb{F}_{64}$, and this is in fact the smallest field having both $\mathbb{F}_4$ and $\mathbb{F}_8$ as subfields.

IX.3 Irreducible polynomials over finite fields

Using the classification of finite fields we deduce some results concerning irreducible polynomials in $\mathbb{F}_q[X]$.

IX.3.1 Example. Suppose K is an extension of $\mathbb{F}_p$ with $[K:\mathbb{F}_p]=2$. Since $\#K=p^2$ and $\#\mathbb{F}_p=p$ there are exactly p^2-p elements in K which are not in $\mathbb{F}_p$. Take one of them: $\alpha\in K$ such that $\alpha\notin\mathbb{F}_p$. Then $\mathbb{F}_p(\alpha)=K$ hence $\deg(f^\alpha_{\mathbb{F}_p})=2$. From Theorem IX.2.4 one knows that $f^\alpha_{\mathbb{F}_p}\in\mathbb{F}_p[X]$ has the two distinct zeros α and α^p in $K-\mathbb{F}_p$. Given K one obtains in this way $\frac{1}{2}(p^2-p)$ irreducible monic polynomials of degree 2 in $\mathbb{F}_p[X]$, and

$$K\cong\mathbb{F}_p[X]/(f)$$

for each of these f.

Vice versa, let $g\in\mathbb{F}_p[X]$ be a monic and irreducible polynomial of degree 2. Then by the classification $\mathbb{F}_p[X]/(g)\cong\mathbb{F}_{p^2}$. Since g is the minimal polynomial over $\mathbb{F}_p$ of $\alpha:=X+(g)\in\mathbb{F}_p[X]/(g)$ it is necessarily one of the $\frac{1}{2}(p^2-p)$ polynomials we found above. Conclusion: There are exactly $\frac{1}{2}(p^2-p)$ monic irreducible polynomials of degree 2 in $\mathbb{F}_p[X]$.

A more direct alternative derivation of the above result runs as follows. A monic polynomial of degree 2 in $\mathbb{F}_p[X]$ has the form X^2+aX+b with $a,b\in\mathbb{F}_p$, so there are p^2 of them. The reducible ones among them are $(X-r)^2$ and $(X-r)(X-s)$ with $r\neq s$ and $r,s\in\mathbb{F}_p$. Hence there are $p+\binom{p}{2}=p+\frac{1}{2}(p^2-p)=\frac{1}{2}(p^2+p)$ reducible ones. We conclude that the number of monic irreducible degree 2 polynomials in $\mathbb{F}_p[X]$ equals $p^2-\frac{1}{2}(p^2+p)=\frac{1}{2}(p^2-p)$, confirming what was found earlier. ∎

IX.3.2 Theorem. *Let $q>1$ be a power of a prime and let $n\in\mathbb{Z}_{\geq 1}$. Then*

$$X^{q^n}-X=\prod f\qquad\text{in }\mathbb{F}_q[X]$$

where the product is taken over the set of monic irreducible polynomials $f\in\mathbb{F}_q[X]$ such that $\deg(f)$ *divides n.*

Proof. Since $\mathbb{F}_q[X]$ is a unique factorization domain, $X^{q^n}-X$ can be factored in a unique way as product of monic irreducible polynomials in $\mathbb{F}_q[X]$. These factors are pairwise distinct since the derivative $(X^{q^n}-X)'=q^nX^{q^n-1}-1=-1$.

The theorem follows once we have shown that

$$f\,\Big|\,X^{q^n}-X\iff\deg(f)\,\Big|\,n,$$

for monic irreducible $f\in\mathbb{F}_q[X]$.

Let $d=\deg(f)$. Since f is monic and $X^{q^n}-X=\prod_{\alpha\in\mathbb{F}_{q^n}}(X-\alpha)\in\mathbb{F}_{q^n}[X]$, we have

$$f\,\Big|\,X^{q^n}-X\iff f=(X-\alpha_1)\cdots(X-\alpha_d)\text{ for some }\alpha_j\in\mathbb{F}_{q^n}[X]\iff\Omega^f_{\mathbb{F}_q}\subseteq\mathbb{F}_{q^n}.$$

Moreover by Theorem IX.2.4 $\Omega^f_{\mathbb{F}_q}=\mathbb{F}_q[\alpha_j]=\mathbb{F}_q[X]/(f)$ with α_j a zero of f (see IX.2.4). Therefore the monic irreducible f of degree d over $\mathbb{F}_q$ satisfies

$$f\,\Big|\,X^{q^n}-X\iff\mathbb{F}_q[X]/(f)\subseteq\mathbb{F}_{q^n}$$

and since $[\mathbb{F}_q[X]/(f):\mathbb{F}_q]=d$ and therefore $\mathbb{F}_q[X]/(f)\cong\mathbb{F}_{q^d}$, Theorem IX.2.7 shows

$$f\,\Big|\,X^{q^n}-X\iff\mathbb{F}_{q^d}\subseteq\mathbb{F}_{q^n}\iff d|n.$$

This finishes the proof. ∎

IX.3.3 Example. Note that

$$X^4 - X = X(X^3 - 1) = X(X+1)(X^2 + X + 1) \qquad \text{in } \mathbb{F}_2[X].$$

Therefore $X, X+1, X^2+X+1$ are the only monic irreducible polynomials of degree ≤ 2 in $\mathbb{F}_2[X]$. Of course this is easy to check directly. ▬

IX.3.4 Corollary. *Let x_d denote the number of monic irreducible degree d polynomials in $\mathbb{F}_q[X]$. For every $n \in \mathbb{Z}_{\geq 1}$ we have*

$$\sum_{d|n} d x_d = q^n.$$

Proof. Observe that $q^n = \deg(X^{q^n} - X)$. By Theorem IX.3.2 the given polynomial equals the product of all monic irreducible polynomials with degree d such that d divides n. The left-hand-side of the formula is precisely the degree of this product. ■

IX.3.5 Remark. Here is an alternative proof of Corollary IX.3.4, which uses only the fact that $\mathbb{F}_q[X]$ is a unique factorization domain. The argument is briefly mentioned in Section 1.1 of the PhD thesis (2008) of the American mathematician Paul Pollack, where he attributes it to his supervisor: the British mathematician Andrew Granville. The details of the argument are as follows.

Denote by $\mathscr{M}$ the set of all monic polynomials in $\mathbb{F}_q[X]$, and by $\mathscr{P} \subset \mathscr{M}$ the subset of *irreducible* monic elements in $\mathbb{F}_q[X]$. Take $n \in \mathbb{Z}_{\geq 1}$ arbitrary, and consider the sum

$$\sum_{\substack{A \in \mathscr{M} \\ \deg(A) = n}} \deg(A).$$

On the one hand, writing such A as $X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0$ with all $a_j \in \mathbb{F}_q$, one observes there are in total q^n such $A \in \mathscr{M}$. They all have degree n, hence the above sum equals nq^n.

Next, one deduces an alternative expression for this sum by factoring each A into monic irreducible elements of $\mathbb{F}_q[X]$, so

$$A = \prod_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } P^a | A}} P.$$

Comparing degrees one obtains

$$nq^n = \sum_{\substack{A \in \mathscr{M} \\ \deg(A) = n}} \sum_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } P^a | A}} \deg(P).$$

Now we interchange the order of summation: considering a fixed $P \in \mathscr{P}$ and $a \geq 1$, we still sum over all $M \in \mathscr{M}$ of degree n such that $P^a | M$. Clearly, if the degree of P^a exceeds n then no such M exist. However for $a\deg(P) = n - b$ with $0 \leq b < n$ every $M = P^a \cdot (X^b + a_{b-1}X^{b-1} + \ldots + a_1 X + a_0)$ works, and there are q^b such extra factors. As a consequence, the above equality can be rewritten as

$$nq^n = \sum_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } a\deg(P) \leq n}} \deg(P) q^{n - a\deg(P)}.$$

Dividing by q^n yields

$$n = \sum_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } a\deg(P) \leq n}} \deg(P) q^{-a\deg(P)}.$$

This formula holds for every $n \geq 1$. Subtracting the formula for the case $n-1$ from the one for n, one obtains

$$1 = \sum_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } a\deg(P)=n}} \deg(P) q^{-a\deg(P)}.$$

Multiplying both sides by $q^n = q^{a\deg(P)}$ this shows

$$q^n = \sum_{\substack{P \in \mathscr{P},\, a \geq 1 \\ \text{such that } a\deg(P)=n}} \deg(P).$$

The summation here is only over those $P \in \mathscr{P}$ such that $d := \deg(P)$ divides n, and given such d the only corresponding a in the summation is $a = n/d$. As before write

$$x_d := \#\{P \in \mathscr{P} : \deg(P) = d\},$$

then by grouping in the above sum the terms with a common degree d one concludes

$$q^n = \sum_{d|n} d x_d,$$

which is exactly Corollary IX.3.4.

IX.3.6 Example. By means of IX.3.4 one obtains a method to recursively determine x_n, the number of irreducible polynomials of degree n over $\mathbb{F}_q$. For $n = 1, 2, 3, 6$ one finds

$$\begin{array}{lllll} 1\cdot x_1 & & & = q^1 & \Rightarrow \quad x_1 = q \\ 1\cdot x_1 & +2\cdot x_2 & & = q^2 & \Rightarrow \quad x_2 = \tfrac{1}{2}(q^2 - q) \\ 1\cdot x_1 & +3\cdot x_3 & & = q^3 & \Rightarrow \quad x_3 = \tfrac{1}{3}(q^3 - q) \\ 1\cdot x_1 & +2\cdot x_2 & +3\cdot x_3 \quad +6\cdot x_6 & = q^6 & \Rightarrow \quad x_6 = \tfrac{1}{6}(q^6 - q^3 - q^2 + q). \end{array}$$

In general one obtains using the *Möbius-inversion-formula* (see Exercise 11 on page 121) that

$$x_n = \frac{1}{n} \sum_{d|n} \mu(d) q^{n/d}, \qquad \text{in which}$$

$$\begin{cases} \mu(n) = 0 & \text{if a prime } p \text{ exists with } p^2 | n; \\ \mu(p_1 p_2 \ldots p_r) = (-1)^r, \end{cases}$$

where the p_i are pairwise distinct primes and $r \in \mathbb{Z}_{\geq 0}$, so in particular $\mu(1) = 1$ (corresponding to $r = 0$). The function μ was introduced in 1831 by the German mathematician and astronomer August Ferdinand Möbius (1790–1868).

The formula for x_n yields in particular

$$n x_n = q^n + \sum_{d|n,\, d>1} \mu(d) q^{n/d} \geq q^n - \sum_{k=1}^{n-1} q^k > 0$$

(the fact that the expression is indeed positive one may show for example by considering how the subtraction is visualized when written out in base q). So this provides a way to verify that indeed irreducible polynomials over $\mathbb{F}_q$ of any degree $n \geq 1$ exist. Using Remark IX.3.5 one therefore obtains a proof of the existence of a finite field with p^n elements, without using theory of splitting fields.

IX.4 The multiplicative group of a finite field

Whereas the *additive* group $(\mathbb{F}_q, +, 0)$ of a finite field with $q = p^n$ and p prime is isomorphic to a product of n copies of $\mathbb{Z}/p/Z$, the *multiplicative* group $\mathbb{F}_q^\times$ is much simpler. For this we recall a result from Chapter III (Corollary III.5.4).

IX.4.1 Theorem. *The multiplicative group $\mathbb{F}_q^\times$ of a finite field $\mathbb{F}_q$ is a cyclic group.*

Proof. Since $\mathbb{F}_q$ is a domain and $\mathbb{F}_q^\times$ is finite, this is a special case of Corollary III.5.4. ∎

IX.4.2 Definition. An element $\alpha \in \mathbb{F}_q^\times$ generating the multiplicative group $\mathbb{F}_q^\times$ is called a *primitive root* of $\mathbb{F}_q$.

So $\alpha \in \mathbb{F}_q^\times$ is a primitive root of $\mathbb{F}_q$ if and only if the order of α in the group $\mathbb{F}_q^\times$ equals $\#\mathbb{F}_q^\times = q-1$. If α is a primitive root of $\mathbb{F}_q$ then every $x \in \mathbb{F}_q^\times$ can be written as

$$x = \alpha^k, \quad \text{and} \quad \Lambda : \mathbb{F}_q^\times \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}, \quad x \mapsto k \bmod (q-1)$$

is an isomorphism of (abelian) groups. In particular k is modulo $q-1$ uniquely determined (by x and α).

IX.4.3 Example. The element $3 \in \mathbb{F}_7$ is a primitive root of $\mathbb{F}_7$ since

$$3^1 = 3, \quad 3^2 = 2, \quad 3^3 = 6, \quad 3^4 = 4, \quad 3^5 = 5, \quad 3^6 = 1$$

and therefore $\mathrm{ord}(3) = 6 = \#\mathbb{F}_7^\times$.

We now construct a primitive root of $\mathbb{F}_9$, which means an element of order 8 in the group $\mathbb{F}_9^\times$. Since the order of any element in a finite group divides the number of elements in the group we only need to find $\alpha \in \mathbb{F}_9^\times$ such that $\alpha^4 \neq 1$.

Note that X^2+1 has no zero in $\mathbb{F}_3$ and therefore

$$\mathbb{F}_9 \cong \mathbb{F}_3[i] := \mathbb{F}_3[X]/(X^2+1), \qquad \text{with} \quad i := X + (X^2+1).$$

Every element of $\mathbb{F}_9$ therefore has a unique representation as $a + bi$ with $a, b \in \mathbb{F}_3$ and moreover $i^2 = -1$. Take $\alpha = 1 + i$, then

$$\alpha^2 = (1+i)^2 = 2i, \quad \alpha^4 = (2i)^2 = -1 \neq 1,$$

hence the order of α equals 8 and α is a primitive root of $\mathbb{F}_9$. You may compute for yourself α^k for $1 \leq k \leq 8$ and check that indeed all elements of $\mathbb{F}_9^\times$ are obtained in this way. ∎

Here as a small application of the existence of a primitive root of $\mathbb{F}_q$ we will provide a (second) proof of the result below. The same result can also be derived by adapting the reasoning used in Example VIII.1.6.

IX.4.4 Corollary. *Take $\alpha \in \mathbb{F}_q^\times$.*
(i). *If $\mathrm{char}(\mathbb{F}_q) = 2$, then α is a square in $\mathbb{F}_q$.*
(ii). *If $\mathrm{char}(\mathbb{F}_q) > 2$, then $m := (q-1)/2 \in \mathbb{Z}$. Now α is a square in $\mathbb{F}_q \;\Leftrightarrow\; \alpha^m = 1$ and $\alpha \in \mathbb{F}_q^\times$ is not a square $\Leftrightarrow\; \alpha^m = -1$.*

Proof. The map 'squaring':

$$\mathbb{F}_q^\times \xrightarrow{x \mapsto x^2} \mathbb{F}_q^\times$$

is a homomorphism with kernel all $x \in \mathbb{F}_q^\times$ such that $x^2 = 1$. If $\mathrm{char}(\mathbb{F}_q) = 2$ then this equation is equivalent to $(x-1)^2 = 0$, which shows that in this case the 'squaring' homomorphism is injective and therefore bijective. This proves (i).

If $\mathrm{char}(\mathbb{F}_q) > 2$ then the 'squaring' map has kernel $\{\pm 1\}$ consisting of two distinct elements. The image of the map is clearly the subgroup of all squares $\mathbb{F}_q^{\times 2} \subset \mathbb{F}_q^\times$. Hence

$$\mathbb{F}_q^\times / \{\pm 1\} \cong \mathbb{F}_q^{\times 2},$$

which shows in particular that the number of squares in $\mathbb{F}_q^\times$ equals $m := (q-1)/2$. By the proof of Theorem IX.1.1

$$\prod_{\alpha \in \mathbb{F}_q} (X - \alpha) = X^q - X = X \cdot (X^{2m} - 1) = X \cdot (X^m - 1) \cdot (X^m + 1).$$

If $\alpha \in \mathbb{F}_q^\times$ is a square, then $\alpha = \gamma^2$ for some $\gamma \in \mathbb{F}_q^\times$ hence $\alpha^m = \gamma^{2m} = 1$. This means that the nonzero squares are zeros of $X^m - 1$. Since their number is m one concludes

$$\prod_{\alpha \in \mathbb{F}_q^{\times 2}} (X - \alpha) = X^m - 1.$$

The $q - m - 1 = m$ non-squares in $\mathbb{F}_q$ are then necessarily the zeros of $X^m + 1$. This finishes the proof.

Here is a slightly different reasoning, showing the same result.
(i): In this case the squares are precisely the image of the Frobenius homomorphism: $\mathbb{F}_q \to \mathbb{F}_q$. Since $\mathbb{F}_q$ is finite, this is an automorphism, from which (i) follows.
(ii): Here $p = \mathrm{char}(\mathbb{F}_q)$ and hence q (which is a power of p) is odd, and therefore $q - 1$ is even and hence $m := (q-1)/2 \in \mathbb{Z}$. Let β be a primitive root of $\mathbb{F}_q$. Then $\mathrm{ord}(\beta) = 2m$ and $\mathbb{F}_q^\times$ consists of the $2m$ pairwise distinct elements

$$\beta, \beta^2, \beta^3, \ldots, \beta^{2m-1}, \beta^{2m} = 1.$$

Here the m elements β^{2k} with $1 \le k \le m$ are clearly squares. Moreover they are zeros of the polynomial $X^m - 1$ since $(\beta^{2k})^m = (\beta^{2m})^k = 1$. So $X^m - 1 = \prod_{k=1}^m (X - \beta^{2k})$. Moreover, if some power, say β^ℓ of β is a square, then there exists an integer n such that $(\beta^n)^2 = \beta^\ell$. Then $2m | (2m - \ell)$ which shows that ℓ is *even*. This shows that the nonzero squares are precisely the zeros of $X^m - 1$.

The *odd* powers of β are *not* zeros of $X^m - 1$; indeed, $(\beta^{2k+1})^m = \beta^{2mk}\beta^m = \beta^m \neq 1$. Also, these elements are not squares. Since

$$0 = \beta^{2m} - 1 = (\beta^m - 1)(\beta^m + 1) \quad \text{and} \quad \beta^m \neq 0,$$

we have $\beta^m = -1$ hence $(\beta^{2k+1})^m = -1$ for all k. One concludes that the zeros of $X^m + 1$ are precisely all non-squares in $\mathbb{F}_q^\times$, as we saw earlier in a slightly different way. ∎

IX.5 Exercises

1. Determine all primitive roots of $\mathbb{F}_7$, $\mathbb{F}_8$, and $\mathbb{F}_9$.
2. Let $p \equiv 3 \bmod 4$ be prime. Show that $\mathbb{Z}[i]/p\mathbb{Z}[i] \cong \mathbb{F}_{p^2}$.
3. Determine $f^{\alpha}_{\mathbb{F}_3}$ for every $\alpha \in \mathbb{F}_9$, and find the factorization of $X^8 - 1$ into irreducible polynomials in $\mathbb{F}_3[X]$.
4. Prove that $\overline{2} + \sqrt{2} \in \mathbb{F}_{25}$ is a primitive root of $\mathbb{F}_{25}$ (here $\sqrt{2}$ is an element of $\mathbb{F}_{25}$ satisfying $\sqrt{2}^2 = \overline{2}$; is it clear why such an element exists?).
5. Prove that $X^4 + \overline{2}$ is irreducible in $\mathbb{F}_{125}[X]$.
6. Show using a counting argument that the number of monic irreducible polynomials in $\mathbb{F}_q[X]$ of degree 3 equals $\frac{1}{3}(q^3 - q)$.
7. Find complete factorizations of the polynomials $X^2 - X$, $X^4 - X$, $X^8 - X$, and $X^{64} - X$ in $\mathbb{F}_2[X]$.
8. Show using Corollary IX.3.4 that
$$\frac{1}{n}q^n \geq x_n \geq \frac{1}{n}(q^n - \frac{q}{q-1}q^{\frac{1}{2}n}).$$
9. Suppose $\mathbb{F}_{p^n} = \mathbb{F}_p(\alpha)$. Show directly that the polynomial $\prod_{i=0}^{n-1}(X - \alpha^{p^i})$ occurring in IX.2.4 has coefficients in $\mathbb{F}_p$ by verifying that each of its coefficients c satisfies $c^p = c$.
10. Let K be a field of characteristic $p > 0$ and suppose $f \in K[X]$ is a polynomial of the form $X^p - X - a$. Let α be a zero of f in some extension field of K.
    (a) Show that $f = \prod_{i \in \mathbb{F}_p}(X - \alpha - i)$ and that $K(\alpha) = \Omega^f_K$.
    (b) Prove that either f is irreducible in $K[X]$, or f factors in $K[X]$ as a product of polynomials of degree 1. (Hint: in case $f = gh$, consider the equality $f(X - j) = g(X - j)h(X - j)$ $(j \in \mathbb{F}_p)$ and conclude that all irreducible factors of f have the same degree.)
    (c) Show that for every $a \in \mathbb{F}_p^{\times}$ the polynomial $X^p - X - a$ is irreducible in $\mathbb{F}_p[X]$.
11. Let R be the ring of arithmetic functions defined in Exercise 27 on page 16. Define $e, E \in R$ by
$$e(n) = \begin{cases} 1 & \text{if} \quad n = 1; \\ 0 & \text{of} \quad n > 1, \end{cases}$$
and
$$E(n) = 1 \text{ for all } n \in \mathbb{Z}_{>0}.$$
By μ we denote the Möbius function introduced in Section IX.3.
    (a) Show that e is the unit element of the ring R.
    (b) Verify that $\mu \in R$.
    (c) Show that $\mu * E = e$ (so μ is the *inverse* of E in $R^{\times}$).
    (d) For $f \in R$ define $g \in R$ by
$$g = F * E, \text{ so } g(n) = \sum_{d|n} f(d) \quad \text{for } n \in \mathbb{Z}_{>0}.$$
    Derive from (c) the *Möbius inversion formula*:
$$f(n) = \sum_{d|n} \mu(d)g(n/d) \quad \text{for } n \in \mathbb{Z}_{>0}.$$
    (e) Prove the formula for x_n presented in Section IX.3.

12. (Compare Exercise 11 on page 100)
    (a) Show that the fields $\mathbb{Q}[X]/(X^2-2)$ and $\mathbb{Q}[Y]/(Y^2-3)$ are not isomorphic.
    (b) Given a prime p we put
    $$R_2 := \mathbb{F}_p[X]/(X^2-\overline{2}) \text{ and } R_3 := \mathbb{F}_p[Y]/(Y^2-\overline{3}).$$
    Determine for all primes p with $2 \le p \le 23$ the structure of these two rings (does the ring contain nilpotent elements, zero divisors, is the ring a field?), and determine for each case whether the two rings are isomorphic.
13. For $q \in \mathbb{Z}$ prime we put $\Phi_q = X^{q-1} + \ldots + X^2 + X + 1 = (X^q - 1)/(X-1) \in \mathbb{Z}[X]$; given p prime we write $f_{q,p} := \Phi_q \bmod p \in \mathbb{F}_p[X]$.
    Take $q = 11$ and p prime and consider
    $$f_{11,p} = g_p := X^{10} + \ldots + X^2 + X + \overline{1} \in \mathbb{F}_p[X].$$
    Prove that all irreducible factors of g_p in $\mathbb{F}_p[X]$ have the same degree.
    Let $G \in \mathbb{F}_p[X]$ be an irreducible factor of g_p. Prove:
    $\deg(G) = 1 \quad \Leftrightarrow \quad (p = 11 \text{ of } p \equiv 1 \pmod{11})$;
    $\deg(G) = 2 \quad \Leftrightarrow \quad p \equiv -1 \pmod{11}$;
    $\deg(G) = 5 \quad \Leftrightarrow \quad p \equiv 3, 4, 5, \text{ or } 9 \pmod{11}$;
    $\deg(G) = 10 \quad \Leftrightarrow \quad p \equiv 2, 6, 7, \text{ or } 8 \pmod{11}$.
14. Take $g = f_{11,3}$ (notation as in Exercise 13), so $g = X^{10} + \ldots + X + \overline{1} \in \mathbb{F}_3[X]$. Find the irreducible factors of g in $\mathbb{F}_3[X]$.
    (Hint: let G be an irreducible factor of g and suppose a is a zero of $G|g$ in an extension of $\mathbb{F}_3$. Show that a^3 and a^9 and $a^{27} = a^5$ and $a^{15} = a^4$ are zeros of G as well. What is the constant term of G?, what are the zeros of H where $g = G \cdot H$?. What are the zeros of $X^5 \cdot G(1/X)$?, which coefficients of G can be determined using the above?).
15. Again we use the notation from Exercise 13.
    (a) Factor $f_{11,5} \in \mathbb{F}_5[X]$.
    (b) Factor $f_{7,13} \in \mathbb{F}_{13}[X]$.
    (c) Factor $f_{13,5} \in \mathbb{F}_5[X]$.
16. (a) Let K be a field and take $x \in K, x^4 \neq 1, x^8 = 1$. Show that $x^4 = -1$ and that $(x + \frac{1}{x})^2 = 2$.
    (b) Determine the order of 3 mod 41 in the group $\mathbb{F}_{41}^\times$. Find $y \in \mathbb{Z}$ such that $y^2 \equiv 2 \bmod 41$.
    (c) Take a prime number $p \equiv 1 \bmod 8$. Show that $z \in \mathbb{Z}$ exists with $z^2 \equiv 2 \bmod p$.
17. (in Exercise 16 we solved the equation $z^2 \equiv 2 \bmod p$ in the case of a prime $p \equiv 1 \bmod 8$. Here we consider $x^2 \equiv 3 \bmod p$ where p is a prime such that $p \equiv 1 \bmod 12$).
    (a) Show that $a \in \mathbb{F}_p^\times$ exists such that $\text{ord}(a) = 12$ in the group $\mathbb{F}_p^\times$.
    (b) Take a as in (a) and let $b = a^2$; show that $b + b^5 = 1$ (hint: verify that $b^3 = -1$ and $(b^2)^2 + b^2 + 1 = 0$).
    (c) For a as above, show that $(a^5 + a^7)^2 = \overline{3} \in \mathbb{F}_p$.
    (d) Prove that $x \in \mathbb{Z}$ exists with $x^2 \equiv 3 \bmod p$.
    (e) Make a sketch of the complex plane and in it $z = e^{2\pi i/12}$ and $z^2 + z^{10}$ and $z^5 + z^7$. Do you see a connection with the other parts of this exercise?
18. Determine the number of irreducible polynomials of degree 4 in $\mathbb{F}_q[X]$.
19. Let $a, b \in \mathbb{F}_p$ $(p > 2)$ such that $X^2 - a$ and $X^2 - b$ are irreducible in $\mathbb{F}_p[X]$.

(a) Prove that $r \in \mathbb{F}_p^\times$ exists with $a = r^2 b$. (Hint: with $\mathbb{F}_p^{\times 2} \subset \mathbb{F}_p^\times$ the subgroup consisting of all squares, consider the group $\mathbb{F}_p^\times / \mathbb{F}_p^{\times 2}$.)

(b) Put $\beta := X + (X^2 - b) \in \mathbb{F}_p[X]/(X^2 - b)$. Show that

$$\mathrm{ev}_{r\beta} : \mathbb{F}_p[X] \longrightarrow \mathbb{F}_p[X]/(X^2 - b), \qquad f \mapsto f(r\beta)$$

is a surjective ring homomorphism with kernel $(X^2 - a)$.

(c) Find an explicit field homomorphism

$$\phi : \mathbb{F}_p[X]/(X^2 - a) \longrightarrow \mathbb{F}_p[X]/(X^2 - b).$$

(d) Let $X^2 + tX + s \in \mathbb{F}_p[X]$ be an irreducible polynomial. Construct a field isomorphism

$$\psi : \mathbb{F}_p[X]/(X^2 + tX + s) \longrightarrow \mathbb{F}_p[X]/(X^2 - b).$$

20. Problem 2 of the 25th International Mathematical Olympiad (1984) reads:

Find one pair of positive integers a, b such that $ab(a+b)$ is not divisible by 7, but $(a+b)^7 - a^7 - b^7$ is divisible by 7^7.

In this exercise some aspects of this problem are discussed.

(a) Let $p \equiv 1 \bmod 3$ be a prime number and take $k \in \mathbb{Z}_{>0}$. Prove that $X^2 + X + 1$ has exactly 2 zeros in $\mathbb{Z}/p^k\mathbb{Z}$. Find these zeros in the cases $p^k = 7^k$, for all $k \in \{1, 2, 3, 4, 5, 6, 7\}$.

(b) Let $p \equiv 1 \bmod 3$ be a prime number. Put $f := \frac{1}{p}((X+1)^p - X^p - 1) \in \mathbb{Q}[X]$. Show that in fact $f \in \mathbb{Z}[X]$, and find irreducible factors a, b, c of f in $\mathbb{Z}[X]$ such that $abc^2 | f$. Hint: take $w = e^{2\pi i/3}$ and compute $(w+1)^6$ (explain the result using a picture of w and $w+1$ in the complex plane!). What is $f(w)$ and what is $f'(w)$?

(c) Write $f = \frac{1}{7}((X+1)^7 - X^7 - 1)$ as a product of monic irreducible polynomials in $\mathbb{Z}[X]$ (This "explains" the first case of Exercise 14 on page 76).

(d) Now solve the IMO problem: find $a, b \in \mathbb{Z}$ satisfying the two conditions

i. $ab(a+b) \not\equiv 0 \pmod 7$,

ii. $(a+b)^7 \equiv a^7 + b^7 \pmod{7^7}$.

(e) How many pairs $(a, b) \in (\mathbb{Z}/7^7\mathbb{Z})^\times \times (\mathbb{Z}/7^7\mathbb{Z})^\times$ exist satisfying the two conditions $a + b$ is a unit and $(a+b)^7 - a^7 - b^7 = \overline{0}$?