Abstract Algebra II

Randall R. Holmes
Auburn University

Last revision: April 25, 2019

Notation

- $\mathbf{N} = \{1, 2, 3 \ldots\}$, natural numbers
- $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, integers
- $\mathbf{Q} = \left\{\frac{m}{n} \,|\, m, n \in \mathbf{Z}, n \neq 0\right\}$, rational numbers (fractions)
- $\mathbf{R}$, real numbers
- $\mathbf{C} = \{a + bi \,|\, a, b \in \mathbf{R}\}$ $(i = \sqrt{-1})$, complex numbers
- $\mathbf{Z}_n = \{0, 1, \ldots, n-1\}$, integers modulo n
- $R^n = \{(r_1, r_2, \ldots, r_n) \,|\, r_i \in R\}$, n-fold cartesian product of the ring R
- R^S, functions from the set S to the ring R
- $R[x]$, polynomials in the indeterminate x with coefficients coming from the ring R
- $\mathrm{End}(A)$, endomorphisms of the abelian group A (i.e., homomorphisms from A to A)
- $\mathbf{H} = \{a + bi + cj + dk \,|\, a, b, c, d \in \mathbf{R}\}$, quaternions
- $\mathrm{Mat}_{m \times n}(\mathbf{R})$, $m \times n$ matrices over $\mathbf{R}$
- $\mathrm{Mat}_n(\mathbf{R})$, $n \times n$ matrices over $\mathbf{R}$

0 Introduction

The general quadratic equation $ax^2 + bx + c = 0$ has solutions given by the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

There are similar formulas for the solutions of the general cubic and quartic equations. For centuries mathematicians tried to find a formula for the solutions of a general quintic equation, but to no avail.

Finally, it was shown by Abel in 1826 that no such formula could possibly exist. He did this by demonstrating that the existence of such a formula would lead to a contradiction, such as $1 = 0$. The same reasoning shows that there is no such formula for the solutions of any polynomial equation of degree greater than four.

The modern-day proof of this theorem involves an area of algebra called Galois theory, named after its main discoverer. Remarkably, the same theory is used to settle other questions that plagued mathematicians for years. For instance, the theory shows that there can be no general algorithm for trisecting an angle using only a straightedge and compass (surprising since there is such an easy algorithm for *bi*secting an angle, which we all learned as children).

We begin by studying general ring theory and then move to field theory and Galois theory. Our goal is to prove, using Galois theory, Abel's result on the insolvability of the quintic (we will prove the nonexistence of an algorithm for trisecting an angle using only straightedge and compass along the way). Aside from the historical significance of this result, the fact that its proof ultimately uses almost every important idea in the course (indeed in both courses) makes it a worthwhile goal.

1 Definition of ring and examples

1.1 Definition

A **ring** is a triple $(R, +, \cdot\,)$, where $(R, +)$ is an abelian group and $\cdot$ is a binary operation on R (written $(r, s) \mapsto rs$) satisfying the following for all $r, s, t \in R$:

(a) $r(st) = (rs)t$,

(b) $r(s + t) = rs + rt$,

(c) $(r + s)t = rt + st$.

If $(R, +, \cdot\,)$ is a ring, we say that R is a ring under $+$ and $\cdot$ (or just that R is a ring when the binary operations are clear from the context). Part (a) says that $\cdot$ is **associative**. Parts (b) and (c) say that $\cdot$ **distributes** over $+$ from the left and the right, respectively.

Let $(R, +, \cdot\,)$ be a ring. Denote by 0 the identity element of the group $(R, +)$.

- An element $1 \neq 0$ of R is an **identity** (or **multiplicative identity**) if it is an identity for the operation $\cdot$, meaning, $1r = r$ and $r1 = r$ for all $r \in R$. An identity, if one exists, is unique. If R has an identity, we sometimes say that R is a ring with 1.

- R is **commutative** if the operation $\cdot$ is commutative, meaning, $rs = sr$ for all $r, s \in R$. R is **noncommutative** if it is not commutative.

1.2 Examples: Z, Q, R, C, 2Z

- $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ are all commutative rings with identity under usual addition and multiplication.

- $2\mathbf{Z} = \{2n \,|\, n \in \mathbf{Z}\}$ is a commutative ring without identity.

1.3 Example: Integers modulo n

Let n be a positive integer and put $\mathbf{Z}_n = \{0, 1, \ldots, n-1\}$. On this set, define **addition modulo** n by letting $r + s$ be the remainder upon division by n of $r + s$ (usual sum). Similarly, define **multiplication modulo** n by letting rs be the remainder upon division by n of rs (usual product). For instance, if $n = 5$, then $4 + 2 = 1$ and $4 \cdot 2 = 3$. Then, with these operations, $\mathbf{Z}_n$ is a ring, the **ring of integers modulo** n. It is commutative, and the number 1 is an identity if $n > 1$.

1.4 Example: R^n

Let R be a ring and let n be a positive integer. The set $R^n = \{(r_1, r_2, \dots, r_n) \,|\, r_i \in R\}$ is a ring under **componentwise** addition and multiplication:

$$\begin{aligned}(r_1, r_2, \dots, r_n) + (s_1, s_2, \dots, s_n) &= (r_1 + s_1, r_2 + s_2, \dots, r_n + s_n),\\ (r_1, r_2, \dots, r_n)(s_1, s_2, \dots, s_n) &= (r_1 s_1, r_2 s_2, \dots, r_n s_n).\end{aligned}$$

If R has identity 1, then the tuple $(1, 1, \dots, 1)$ is an identity for R^n.

1.5 Example: Functions into a ring

Let S be a nonempty set, let R be a ring, and let R^S denote the set of all functions from S to R. For $f, g \in R^S$, define $f + g$ and fg in R^S by $(f + g)(s) = f(s) + g(s)$ and $(fg)(s) = f(s)g(s)$ $(s \in S)$. R^S is a ring under this addition and multiplication. It is commutative if and only if R is commutative. If R has identity 1, then R^S has identity, also denoted 1, defined by $1(s) = 1$ for all $s \in S$.

If $S = \{1, 2, \dots, n\}$, then we can identify $f \in R^S$ with the n-tuple $(r_1, r_2, \dots, r_n)$, where $r_i = f(i)$ $(1 \leq i \leq n)$, and thereby identify the ring R^S with the ring R^n. It is because of the terminology in this special case that one often refers to the operations defined above for general S as componentwise addition and multiplication.

1.6 Example: Polynomial ring

Let R be a ring. A **polynomial** over R in the **indeterminate** x, is an expression of the form

$$\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n,$$

with n a nonnegative integer and each **coefficient** a_i an element of R.

In the polynomial above, if $i > n$ we put $a_i = 0$ so that a_i is defined for every nonnegative integer i.

Two polynomials are equal if and only if their corresponding coefficients are equal:

$$\sum_i a_i x^i = \sum_i b_i x^i \qquad \Longleftrightarrow \qquad a_i = b_i \text{ for all } i.$$

Polynomials are added and multiplied using the usual rules:

$$\sum_i a_i x^i + \sum_i b_i x^i = \sum_i (a_i + b_i) x^i,$$
$$(\sum_i a_i x^i) \cdot (\sum_i b_i x^i) = \sum_i c_i x^i,$$

where $c_i = \sum_{j=0}^{i} a_j b_{i-j}$. With these operations,

$$R[x] := \{\sum_{i=0}^{n} a_i x^i \,|\, n \in \mathbf{N} \cup \{0\}, a_i \in R\}$$

is a ring, the **polynomial ring** over R in the indeterminate x.

The definition of $R[x]$ given here lacks rigor (for instance, "expression" is a vague term), but it conforms to our earlier experiences with polynomials and it is suitable for our discussions here. The reader interested in a careful definition can find one in Section 9.1.

1.7 Example: Matrix ring

Let R be a ring and let n be a positive integer. Denote by $\mathrm{Mat}_n(R)$ the set of all $n \times n$ matrices with entries coming from R. This is a ring under matrix addition and matrix multiplication (carried out using the operations in R), the **matrix ring** of degree n over R. It is noncommutative if R has an identity and $n > 1$, and also if R is noncommutative. If R has identity 1 then the usual identity matrix I (having 1's down the main diagonal and 0's elsewhere) is an identity for $\mathrm{Mat}_n(R)$.

1.8 Example: Endomorphism ring

Let A be an abelian group (with binary operation +). Define

$$\mathrm{End}(A) = \{f : A \to A \,|\, f \text{ is a group homomorphism}\}.$$

Let $f, g \in \mathrm{End}(A)$. Define $f + g : A \to A$ by $(f + g)(a) = f(a) + g(a)$ and $f \circ g : A \to A$ by $(f \circ g)(a) = f(g(a))$ (function composition). Then $f + g$ and $f \circ g$ are both elements of $\mathrm{End}(A)$.

$(\mathrm{End}(A), +, \circ)$ is a ring, the **endomorphism ring** of A. It has identity $1 = 1_A$ defined by $1_A(a) = a$ for all $a \in A$.

1.9 Example: Quaternion ring

Let $\mathbf{H}$ be the set of expressions of the form $a+bi+cj+dk$ with $a, b, c, d \in \mathbf{R}$ (like complex numbers $a+bi$ with two more terms). View such expressions as polynomials in the indeterminates i, j, and k. Define addition in $\mathbf{H}$ to be the same as polynomial addition,

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

and define multiplication in $\mathbf{H}$ to be the same as polynomial multiplication except subject to the rules

$$\begin{array}{llll} i^2, j^2, k^2 = -1, & ij = k, & jk = i, & ki = j, \\ & ji = -k, & kj = -i, & ik = -j, \end{array}$$

so that, for instance,

$$\begin{aligned} (2+3i-k)(4-j+5k) &= 8-2j+10k+12i-3ij+15ik-4k+kj-5k^2 \\ &= 8-2j+10k+12i-3k-15j-4k-i+5 \\ &= 13+11i-17j+3k. \end{aligned}$$

(An easy way to remember the last six rules is to think of i, j, and k as the standard unit vectors in $\mathbf{R}^3$ and multiply them using the cross product, noting that the signs are determined by the right-hand rule.)

The set $\mathbf{H}$, with addition and multiplication as just described, is a ring, the **quaternion ring**. Checking that the ring axioms are satisfied is straightforward but tedious (but see Section 2.10).

1 – Exercises

1–1 Let S be a nonempty set and let R be a ring. Verify that the left distributive law ((b) of 1.1) holds in R^S (see 1.5).

HINT: The left distributive law states that $f(g+h) = fg + fh$ for all $f, g, h \in R^S$. Each side of this equation represents a function. Two functions $F, G : S \to R$ are equal if and only if $F(s) = G(s)$ for all $s \in S$.

1–2 Let A be an abelian group. Verify that the left distributive law holds in $\mathrm{End}(A)$ (see 1.8).

1–3 Let R be a ring. Give an inductive proof of the **generalized left distributive law**: $r(s_1 + s_2 + \cdots + s_n) = rs_1 + rs_2 + \cdots + rs_n$ for all $r, s_i \in R$.

1–4 Let R be a ring and assume that $r^2 = r$ for all $r \in R$.

(a) Prove that $r + r = 0$ for all $r \in R$.

(b) Prove that R is commutative.

2 Elementary notions

2.1 Notation in underlying abelian group

Let $(R, +, \cdot\,)$ be a ring. The group $(R, +)$ is the **underlying abelian group** of the ring R. The usual notational conventions are used for this group:

- The identity of $(R, +)$ is denoted 0. It is called the **additive identity** to distinguish it from a possible multiplicative identity.
- The inverse of an element r of the group $(R, +)$ is denoted $-r$. It is called the **additive inverse** of r to distinguish it from a possible multiplicative inverse.
- Let $r \in R$. For an integer n, the expression nr has the meaning that it was given in group theory: define $0r = 0$, where the 0 on the left is the integer, and the 0 on the right is the additive identity of R; for a positive integer n, define $nr = r + r + \cdots + r$ (n summands), and $(-n)r = n(-r)$ (which equals $-nr$).

If $n \in \mathbf{Z}$ and $r \in R$, the expression nr has two interpretations if n happens to be an element of R, namely, as defined above and also as a product of two ring elements. In common rings for which this can happen (like $\mathbf{Z}$, $\mathbf{Q}$, and $\mathbf{R}$), the two possible interpretations of the expression coincide.

The trivial group $\{0\}$ is a ring (with multiplication given by $0 \cdot 0 = 0$). It is the **trivial ring**. Although 0 is an identity for the multiplication in this ring, the ring does *not* have an identity, according to the definition, because it is required that an identity be nonzero (see Section 1.1). The reason for this choice is that there are several statements that hold for rings with identity but not for the trivial ring and it keeps us from constantly having to make this exception.

2.2 Basic identities

Let R be a ring and let $r, s \in R$.

2.2.1 Theorem.

(i) $0r = 0$ *and* $r0 = 0$*, where* 0 *is the additive identity of* R*,*

(ii) $r(-s) = -rs$ *and* $(-r)s = -rs$*,*

(iii) $(-r)(-s) = rs$*.*

Proof. (i) Using the right distributive property, we get

$$0r + 0r = (0 + 0)r = 0r = 0r + 0,$$

so left cancellation gives $0r = 0$. Similarly, $r0 = 0$.

(ii) Since $rs + r(-s) = r(s + (-s)) = r0 = 0$, using part (i), it follows that $r(-s) = -rs$. Similarly, $(-r)s = -rs$.

(iii) Using part (ii) twice, we get $(-r)(-s) = -(-r)s = -(-rs) = rs$. □

2.3 Characteristic of a ring with identity

Let R be a ring with identity. The **characteristic** of R, denoted $\mathrm{char}(R)$, is the order of the element 1 in the underlying abelian group $(R, +)$ unless this order is infinity, in which case $\mathrm{char}(R) = 0$. Thus, if there exists a positive integer n such that $1 + 1 + \cdots + 1 = 0$ (n summands), then the least such n is the characteristic of R. Otherwise, R has characteristic zero.

For each integer $n > 1$, the ring $\mathbf{Z}_n$ has characteristic n. The rings $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ each have characteristic zero.

Because of the requirement $1 \neq 0$ it follows that if R has nonzero characteristic n, then $n \neq 1$, implying $n > 1$.

2.4 Unit in a ring with identity

Let R be a ring with identity. An element r of R is a **unit** if it has a multiplicative inverse, that is, if there exists s in R such that $rs = 1$ and $sr = 1$. In this case, s is uniquely determined by r; it is denoted r^{-1}. A **nonunit** is an element that is not a unit. Note that $0 \in R$ is a nonunit since, for every $s \in R$ we have $0s = 0 \neq 1$.

The set $U(R)$ of all units in R is a group under multiplication, the **group of units** of R.

For example, $U(\mathbf{Z}) = \{1, -1\} \cong \mathbf{Z}_2$.

2.5 Units in $\mathbf{Z}_n$

The **Euler phi function** is the function $\phi : \mathbf{N} \to \mathbf{N} \cup \{0\}$ with $\phi(n)$ defined to be the number of integers m such that $1 \leq m < n$ and $\gcd(m, n) = 1$.

For example, $\phi(9) = 6$ since 1, 2, 4, 5, 7 and 8 are the integers satisfying the conditions on m in this case.

Let n be an integer with $n > 1$.

2.5.1 Theorem.

(i) *A nonzero element m of $\mathbf{Z}_n$ is a unit if and only if* $\gcd(m, n) = 1$.

(ii) *$U(\mathbf{Z}_n)$ has order $\phi(n)$.*

(iii) *$U(\mathbf{Z}_p) = \mathbf{Z}_p \backslash \{0\}$ for each prime number p.*

Proof. (i) Let m be a nonzero element of $\mathbf{Z}_n$. The set $H = \{am + bn \mid a, b \in \mathbf{Z}\}$ is a subgroup of $\mathbf{Z}$ and it is cyclic since $\mathbf{Z}$ is cyclic. In group theory, $\gcd(m, n)$ was defined to be the positive generator of H (and it was shown that this definition coincides with the usual one).

Now m is a unit if and only if $rm = 1$ for some $r \in \mathbf{Z}_n$, the product rm being taken modulo n. Therefore, m is a unit if and only if $rm = sn + 1$ for some $r, s \in \mathbf{Z}$, where here the computations are carried out in $\mathbf{Z}$. Since this equation can be written $1 = rm + (-s)n$, we see that m is a unit if and only if 1 is an element of H, which holds if and only if $H = \langle 1 \rangle$. The claim follows.

(ii) This follows immediately from part (i) and the definition of $\phi(n)$.

(iii) If p is a prime number, then $\gcd(m, p) = 1$ for all $0 < m < p$, so the claim follows from (i). □

For example, the units in $\mathbf{Z}_9$ are 1, 2, 4, 5, 7 and 8. Therefore, $U(\mathbf{Z}_9) = \{1, 2, 4, 5, 7, 8\}$, a group of order $\phi(9) = 6$ (and hence isomorphic to $\mathbf{Z}_6$ since it is abelian).

2.6 Division ring and field

A **division ring** is a ring with identity having the property that every nonzero element is a unit. Put another way, a ring R with identity is a division ring if and only if $U(R) = R^\times$, where $\times$ indicates that 0 is removed (so $R^\times = R\backslash\{0\}$).

A **field** is a commutative division ring.

- $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ are fields.
- $\mathbf{Z}_p$ is a field for every prime number p (see Section 2.5).
- $\mathbf{H}$ is a division ring but not a field (see Exercise 2–7).
- $\mathbf{Z}$ is not a division ring since $2 \in \mathbf{Z}$ has no multiplicative inverse.

2.7 Direct sum of rings

Let R_1 and R_2 be rings. The direct sum $R_1 \oplus R_2 = \{(r_1, r_2) \,|\, r_i \in R_i\}$ of the underlying abelian groups of R_1 and R_2 is a ring with componentwise multiplication. Therefore, the operations in this ring are given by

$$(r_1, r_2) + (r_1', r_2') = (r_1 + r_1', r_2 + r_2'), \qquad (r_1, r_2)(r_1', r_2') = (r_1 r_1', r_2 r_2').$$

This ring is the **direct sum** of the rings R_1 and R_2. The direct sum of an arbitrary finite collection of rings is defined similarly.

2.8 Operation tables

Let R be a finite ring. There are two operation tables associated with R: the **addition table** of R is the operation table of the underlying abelian group described in group theory; the **multiplication table** of R, defined analogously, is the table with rows and columns labeled with the elements of R (in a fixed order, usually with 0 coming first) and with the product rs displayed in the row labeled r and the column labeled s $(r, s \in R)$.

For example, the ring $\mathbf{Z}_4$ has addition and multiplication tables

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

The addition table of R has the property that each element of the ring appears precisely once in each row and in each column. The multiplication table of a nontrivial ring R does not satisfy this property due to the fact that $0r = 0$ for all $r \in R$. However, if R is a division ring, then the multiplication table with 0's removed does satisfy this property since $(R^\times, \,\cdot\,)$ is a group.

2.9 Isomorphism

Let R and R' be rings. An **isomorphism** from R to R' is a bijection $\varphi : R \to R'$ satisfying the **homomorphism property**:

(i) $\varphi(r + s) = \varphi(r) + \varphi(s)$,

(ii) $\varphi(rs) = \varphi(r)\varphi(s)$,

for all $r, s \in R$. So $\varphi : R \to R'$ is an isomorphism if it is an isomorphism between the underlying groups and it also satisfies the multiplicative property stated in (ii).

The rings R and R' are **isomorphic**, written $R \cong R'$, if there exists an isomorphism from R to R'. (This definition appears to be asymmetrical, but in fact $R \cong R'$ if and only if $R' \cong R$ by Exercise 2–5.)

Assume $R \cong R'$ and let $\varphi : R \to R'$ be an isomorphism. As with the notion of group isomorphism, φ can be viewed as a "renaming function"; it takes an element r of R and renames it $\varphi(r)$ in R'. Since φ is injective, no two elements of R end up with the same name after renaming, and since it is surjective, every name in R' gets used. Moreover, since φ satisfies the homomorphism property, the binary operations in R' act on the renamed elements in exactly the same way the binary operations in R act on the elements before renaming. Consequently, the rings R and R' are exactly the same except possibly for the names chosen for their elements and the symbols chosen for their binary operations.

2.10 Example

Since two isomorphic rings are identical, except possibly for the notation used for their elements and the symbols used for their binary operations (Section 2.9), it follows that if one has a property expressible entirely in terms of its elements and its binary operations, then the other must also have that property. Here is an illustration of this principle:

2.10.1 Theorem. *Let R and R' be rings and assume that $R \cong R'$. If R is commutative, then so is R'.*

Proof. Assume R is commutative. Let r' and s' be two elements of R'. Since $R \cong R'$, there exists an isomorphism $\varphi : R \to R'$. In particular, φ is surjective, so there exist r and s in R such that $\varphi(r) = r'$ and $\varphi(s) = s'$. We have

$$r's' = \varphi(r)\varphi(s) = \varphi(rs) = \varphi(sr) = \varphi(s)\varphi(r) = s'r'.$$

Therefore, R' is commutative. □

If you know that a given ring has a certain property and you wish to conclude that a ring isomorphic to it has the same property, then it is usually safe to skip this formalism of a theorem and proof and simply draw the conclusion. For instance, assuming a ring R is isomorphic to a ring R', it is customary just to assert: if R has an identity, then so does R'; if R has characteristic 5, then so does R'; if R has precisely three nonunits, then so does R'; and so on.

The definition of isomorphism assumes that R and R' are both known to be rings, and this will usually be the case in the applications of this notion. However, the definition makes sense even if it is assumed only that these are sets with two binary operations. We give an example to show that there is sometimes something to be gained by relaxing the assumptions.

It was mentioned that it is tedious to check that the addition and multiplication defined for the set $\mathbf{H}$ satisfy the ring axioms (see 1.9). Here is a way to avoid some of the computations. Let R be the set of 2×2 matrices of the form

$$\begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}.$$

Then R is closed under matrix addition and multiplication (as is easy to check) and is therefore a ring (a "subring" of $\mathrm{Mat}_2(\mathbf{C})$). Define $\varphi : \mathbf{H} \to R$ by

$$\varphi(a+bi+cj+dk) = \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}.$$

Then φ is a bijection and it satisfies the homomorphism property (see Exercise 2–6). Therefore, this map is simultaneously an isomorphism of additive binary structures and multiplicative binary structures: $(\mathbf{H}, +) \to (R, +)$, $(\mathbf{H}, \cdot\,) \to (R, \cdot\,)$. Since addition and multiplication in R satisfy the ring axioms, it follows that addition and multiplication in $\mathbf{H}$ satisfy the ring axioms as well.

2 – Exercises

2–1 Let R be a ring. Prove that $r(ns) = n(rs)$ for all $r, s \in R$, $n \in \mathbf{Z}$.

Hint: First handle the case of positive n. Use Section 2.2 for the remaining cases.

2–2 Let R be a ring with identity 1 and let $n \in \mathbf{Z}$. Prove that $nr = 0$ for every $r \in R$ if and only if $n1 = 0$. (Note: This shows, in particular, that $nr = 0$ for every $r \in R$, where $n = \mathrm{char}(R)$.)

2–3 Find the characteristic of the ring $\mathrm{End}(A)$, where $A = \mathbf{Z}_4 \oplus \mathbf{Z}_6$.

2–4 Determine $|U(\mathbf{Z}_8 \oplus \mathbf{Z}_{12})|$ and support your claim.

2–5 Prove that the property of being isomorphic ($\cong$) is an equivalence relation (i.e., reflexive, symmetric, transitive) on the class of all rings. (You may skip the proof of any part involving the underlying abelian group since the corresponding statement about groups is known to be valid.)

2–6 Verify that $\varphi : \mathbf{H} \to R$ as defined in Section 2.10 satisfies the homomorphism property.

2–7 Prove that the ring $\mathbf{H}$ of quaternions is a division ring but not a field.

Hint: Due to Exercise 2–6 and the discussion in Section 2.10 we know that $\mathbf{H}$ is a ring. $\mathbf{H}$ can be viewed as a generalization of $\mathbf{C}$. A formula for the inverse of a nonzero complex number $z = a + bi$ is $z^{-1} = \bar{z}/|z|^2$, where $\bar{z} = a - bi$ and $|z| = \sqrt{a^2 + b^2}$.

3 Subring and ideal

3.1 Definition

Let R be a ring. A subset S of R is a **subring** of R, written $S \leq R$ if it is a subgroup of $(R, +)$ and it is closed under multiplication. Therefore, a subset S of R is a subring if

(i) $0 \in S$,

(ii) $s, s' \in S \Rightarrow s + s' \in S$,

(iii) $s \in S \Rightarrow -s \in S$,

(iv) $s, s' \in S \Rightarrow ss' \in S$.

(The first three properties say that S is a subgroup of $(R, +)$.) A subring is a ring in its own right.

A subring I of R is an **ideal** of R, written $I \triangleleft R$, if it satisfies the **absorption property**

$$a \in I, r \in R \quad \Rightarrow \quad ra, ar \in I.$$

An ideal is the ring-theoretic analog of group theory's normal subgroup. The set of cosets of an ideal form a ring using natural operations (see Section 5).

The absorption property is stronger than the statement that I is closed under multiplication. Therefore, in order to show that a subset I of R is an ideal, one need only show that it is a subgroup of $(R, +)$ and that it satisfies the absorption property.

There are one-sided versions of an ideal: A **left ideal** of R is a subring I that satisfies the left absorption property ($a \in I, r \in R \Rightarrow ra \in I$), and similarly for **right ideal**. In order to distinguish from either of these versions, an ideal, as defined above, is often referred to as a **two-sided ideal**.

3.2 Examples

- Let R be a ring. Then R is an ideal of R; any other ideal is a **proper ideal**. Also $\{0\}$ is an ideal of R, the **trivial ideal**; any other ideal is a **nontrivial ideal**.
- If n is an integer, then $n\mathbf{Z}$ is an ideal of $\mathbf{Z}$.
- We have nested subrings $\mathbf{Z} \leq \mathbf{Q} \leq \mathbf{R} \leq \mathbf{C} \leq \mathbf{H}$ none of which is an ideal of the next.
- The set $S = \{(n, 0) \,|\, n \in \mathbf{Z}\}$ is a subring of $\mathbf{Z} \oplus \mathbf{Z}$ (it is even an ideal). This example shows that if a ring has an identity, then a subring need not have that identity as an element. Indeed $\mathbf{Z} \oplus \mathbf{Z}$ has identity $(1, 1)$, but $(1, 1) \notin S$. The subring S does happen to have its own identity though, namely, $(1, 0)$.

3.3 Ideal containing unit is whole ring

Let R be a ring with identity and let I be an ideal of R.

3.3.1 Theorem. *If I contains a unit, then $I = R$.*

Proof. Assume that I contains a unit u. For every $r \in R$, we have

$$r = r1 = r(u^{-1}u) = (ru^{-1})u \in I,$$

where we have used the absorption property of I. Therefore, $R \subseteq I$, which forces equality. □

In particular, the only ideals in a field F are $\{0\}$ and F.

3.4 Ideal generated by a set

Let R be a ring and let X be a subset of R. The **ideal of R generated by X**, denoted (X), is the intersection of all ideals of R containing X:

$$(X) = \bigcap_{\substack{I \lhd R \\ I \supseteq X}} I.$$

Since the intersection of a collection of ideals is an ideal (Exercise 3–4), (X) is indeed an ideal of R as the terminology suggests. It is the smallest ideal containing X in the sense that it is contained in every ideal that contains X.

If $X = \{a_1, a_2, \ldots, a_n\}$, then (X) equals $(\{a_1, a_2, \ldots, a_n\})$, but we write this more simply as $(a_1, a_2, \ldots, a_n)$.

3.5 Principal ideal

Let R be a ring. For $a \in R$, the ideal (a) is the **principal ideal** of R generated by a.

3.5.1 Theorem. *Assume that R is commutative and that it has an identity. For every $a \in R$ we have*

$$(a) = Ra := \{ra \,|\, r \in R\}.$$

Proof. Let $a \in R$. We claim that Ra is an ideal. First, $0 = 0a \in Ra$. Next, for each $r, r' \in R$ we have $ra + r'a = (r + r')a \in Ra$, and $-(ra) = (-r)a \in Ra$ so Ra is closed under addition and negation. Finally, for each $r, s \in R$ we have $r(sa) = (rs)a \in Ra$ and $(sa)r = (sr)a \in Ra$ (using that R is commutative), so the absorption property holds.

Now $a = 1a \in Ra$, so $(a) \subseteq Ra$ since (a) is the intersection of all ideals of R containing a. On the other hand, every ideal of R containing a must contain Ra by the absorption property, so that $Ra \subseteq (a)$. Therefore, $(a) = Ra$ as claimed. □

Here are some examples of principal ideals.

- In the ring $\mathbf{Z}$, we have $(n) = n\mathbf{Z}$ for each $n \in \mathbf{Z}$. Since an ideal of $\mathbf{Z}$ is a subgroup of $(\mathbf{Z}, +)$ and therefore of the form $n\mathbf{Z}$ for some $n \in \mathbf{Z}$, it follows that every ideal of $\mathbf{Z}$ is principal.

- In the ring $\mathbf{Z}[x]$, the principal ideal (x) consists of all polynomials having no constant term.
- In the ring $\mathbf{Z} \oplus \mathbf{Z}$, we have $((1,0)) = \{(n,0) \,|\, n \in \mathbf{Z}\}$.

3 – Exercises

3–1 Let R be a ring. The **center** of R is the set

$$C = \{c \in R \,|\, cr = rc \text{ for all } r \in R\}.$$

Prove that the center of R is a subring of R.

3–2 Let R be a commutative ring with identity and let n be a positive integer. Prove that the center of $\mathrm{Mat}_n(R)$ (see Exercise 3–1) is the set of scalar matrices, where a **scalar matrix** is a matrix of the form rI $(r \in R)$ (i.e., r's on the main diagonal and zeros elsewhere).

HINT: Any matrix in the center must commute with each matrix e_{kl} having 1 in the (k,l) position and zeros elsewhere. Use that the (i,j) entry of e_{kl} is $\delta_{ik}\delta_{jl}$, where δ_{ab} is the **Kronecker delta** defined to be 1 if $a = b$ and zero otherwise.

3–3 Let R be a ring and let I and J be ideals of R. Define the **sum** of I and J by $I + J = \{a + b \mid a \in I, b \in J\}$. Prove that $I + J$ is an ideal of R.

3–4 Let R be a ring and let $\{I_\alpha\}_{\alpha \in A}$ be a family of ideals of R. Prove that the intersection $\bigcap_{\alpha \in A} I_\alpha$ is an ideal of R.

4 Integral domain

4.1 Definition

A nonzero element r of a ring is a **divisor of zero** if there exists a nonzero element s of the ring such that $rs = 0$. An **integral domain** is a commutative ring with identity having no divisors of zero. Put another way, a ring is an integral domain if and only if it is a commutative ring with identity and for elements r and s of the ring,

$$rs = 0 \quad \Rightarrow \quad r = 0 \text{ or } s = 0.$$

- $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ are integral domains. In fact, every field is an integral domain (see Section 4.3).
- $\mathbf{Z}$ is an integral domain.
- $\mathbf{Z}_6$ is not an integral domain, since 2 and 3 are nonzero elements having product zero. More generally, $\mathbf{Z}_n$ $(n > 1)$ is an integral domain if and only if n is prime.

4.2 Cancellation property

Let R be a commutative ring with identity. The ring R is said to have the **cancellation property** if, for $r, s, t \in R$,

$$(rs = rt \text{ and } r \neq 0) \quad \Rightarrow \quad s = t.$$

4.2.1 Theorem. *R is an integral domain if and only if it has the cancellation property.*

Proof. Assume that R is an integral domain. Let $r, s, t \in R$ with $rs = rt$ and $r \neq 0$. Rearranging, we have $r(s - t) = 0$. The assumption gives $s - t = 0$, so that $s = t$. Therefore, R has the cancellation property.

Now assume that R has the cancellation property. We have assumed that R is a commutative ring with identity. Let $r, s \in R$ with $rs = 0$. Assume that $r \neq 0$. We have $rs = 0 = r0$, so the cancellation property yields $s = 0$. Therefore, R is an integral domain. □

4.3 Every field is an integral domain

4.3.1 Theorem. *Every field is an integral domain.*

Proof. Let R be a field. As part of the definition of field, R is a commutative ring with identity. The zero element of R is not a divisor of zero (by one of the requirements of divisor of zero) and every nonzero element of R is a unit and hence not a divisor of zero (by Exercise 4–1). Thus, R has no divisors of zero and is therefore an integral domain. □

The converse of this theorem is not true since the ring of integers $\mathbf{Z}$ is an integral domain but not a field. However, we see in the next section that a *finite* integral domain is a field.

4.4 Every finite integral domain is a field

Let n be a positive integer. If n is composite, say, $n = lm$ with $l, m \in \mathbf{N}$, $l, m < n$, then l is a divisor of zero in $\mathbf{Z}_n$, since $lm = 0$. It follows that the only integral domains among the rings $\mathbf{Z}_n$, $n = 2, 3, \ldots$, are the ones with n prime, that is, the fields. This is in keeping with the following result.

4.4.1 Theorem. *Every finite integral domain is a field.*

Proof. Let R be a finite integral domain and let r be a nonzero element of R. It suffices to show that r is a unit. Since R is finite, its elements can be enumerated: $r_1, r_2, \ldots, r_n$. Since R has the cancellation property (Section 4.2), the elements $rr_1, rr_2, \ldots, rr_n$ are distinct. Therefore, these must be the n elements of R in some order. In particular, $rr_i = 1$ for some i, implying that r is a unit. □

4.5 Characteristic of integral domain

Let R be an integral domain.

4.5.1 Theorem.

(i) *The characteristic of R is either prime or zero.*

(ii) *If R is finite, then its characteristic is prime.*

Proof. (i) Assume that the characteristic n of R is not zero. First note that we have $n > 1$. Suppose that n has a proper factorization: $n = n_1 n_2$ with $0 < n_1, n_2 < n$. Using Exercise 2–1, we get $(n_1 1)(n_2 1) = n1 = 0$, and since R is an integral domain, it follows that either $n_1 1 = 0$ or $n_2 1 = 0$. But either case contradicts that n is the order of 1 in $(R, +)$. We conclude that n has no proper factorization, that is, $\mathrm{char}(R) = n$ is prime.

(ii) We prove the contrapositive. Assume that the characteristic of R is not prime. By (i), the characteristic of R is zero, which is to say that 1 has infinite order in the group $(R, +)$. Thus $\mathbf{Z} \cong \langle 1 \rangle \subseteq R$ and R is infinite. $\square$

4.6 Field of fractions of an integral domain

The ring $\mathbf{Z}$ is an integral domain. However, it is not a field, since, for instance, the integer 2 is not a unit (it has no multiplicative inverse). One imagines that $\mathbf{Z}$ is not a field because it is just not large enough.

The question arises of whether the set $\mathbf{Z}$ can be augmented by the addition of new elements in such a way that the resulting set can be made into a field with operations compatible with those in $\mathbf{Z}$. It is these considerations that led to the creation of the field $\mathbf{Q}$, which is the set of all fractions n/m $(n, m \in \mathbf{Z},\ m \neq 0)$ endowed with usual fraction addition and multiplication. The ring $\mathbf{Z}$ is viewed as a subring of $\mathbf{Q}$ by identifying the integer n with the fraction $n/1$.

There are other fields that contain $\mathbf{Z}$ as a subring, but $\mathbf{Q}$ is the smallest such in the sense that any other contains $\mathbf{Q}$ as well (or at least a field isomorphic to $\mathbf{Q}$). For instance, $\mathbf{R}$ and $\mathbf{C}$ both contain $\mathbf{Z}$ as a subring, and they contain $\mathbf{Q}$ as well.

The properties of an integral domain are strong enough to guarantee that the procedure described above can be carried out almost verbatim to embed the integral domain in a field. Here is the construction:

Let R be an integral domain.

(i) Let P be the set of all pairs (r, s) with $r, s \in R$ and $s \neq 0$. (Think of (r, s) as the fraction r/s.)

(ii) Define addition and multiplication on P by the following formulas:

$$(r, s) + (r', s') = (rs' + sr', ss'), \qquad (r, s)(r', s') = (rr', ss').$$

(If P is to be closed under these operations, we must have $ss' \neq 0$, and this is the case since R has no divisors of zero.)

(iii) Define a relation on P by putting

$$(r, s) \sim (r', s') \text{ if } rs' = sr'$$

This is an equivalence relation. Denote by r/s the equivalence class of (r, s). (Thus, just like with rational numbers, we have equality of fractions, $\frac{r}{s} = \frac{r'}{s'}$, if and only if the “cross products” rs' and sr' are equal.)

(iv) The addition and multiplication defined on P induce operations on the set $Q(R) = \{r/s \,|\, r, s \in R, s \neq 0\}$:

$$\frac{r}{s} + \frac{r'}{s'} = \frac{rs' + sr'}{ss'}, \qquad \frac{r}{s} \cdot \frac{r'}{s'} = \frac{rr'}{ss'},$$

and with these operations, $Q(R)$ is a field.

(v) The set $\{r/1 \,|\, r \in R\}$ is a subring of $Q(R)$. The map defined by $r \mapsto r/1$ is an isomorphism from R onto this set. We use this isomorphism to view R as a subring of $Q(R)$ (just as we view $\mathbf{Z}$ as a subring of $\mathbf{Q}$ by writing the integer n as the fraction $n/1$).

(vi) If F is a field and R is a subring of F, then $\{rs^{-1} \,|\, r, s \in R, s \neq 0\}$ is a subring of F isomorphic to $Q(R)$ and it contains R. (So $Q(R)$ is the smallest field containing R in a sense.)

4 – Exercises

4–1 Let R be a commutative ring with identity and let r be a unit in R. Prove that r is not a divisor of zero.

4–2 Let R be a commutative ring with identity and assume that $\{0\}$ and R are the only ideals of R. Prove that R is a field.

Hint: Section 3.5.

4–3 Let R be a ring having no divisors of zero and let I and J be nontrivial ideals of R. Prove that $I \cap J$ is nontrivial.

4–4 Prove that the addition defined in (iv) of 4.6 is well defined in the sense that it does not depend on how the fractions are written.

4–5 Prove (vi) of 4.6.

5 Quotient ring

5.1 Definition

Let R be a ring and let I be an ideal of R. Since the group $(R, +)$ is abelian, the ideal I, viewed as a subgroup, is normal. Therefore, the quotient group R/I is defined. It is the set

$$R/I = \{r + I \,|\, r \in R\}$$

of cosets of I with binary operation being coset sum. The next theorem says that there is a product of cosets relative to which R/I is actually a ring.

5.1.1 Theorem.

(i) *The formulas*

$$(r + I) + (s + I) = (r + s) + I \quad \textit{and} \quad (r + I)(s + I) = rs + I.$$

give well-defined binary operations $+$ *and* $\cdot$ *on* R/I.

(ii) $(R/I, +, \cdot\,)$ *is a ring.*

Proof. (i) In group theory it is shown that the indicated sum is well-defined. It is also shown in group theory that there is equality of cosets $r+I = r'+I$ if and only if $r - r' \in I$. Let $r, r', s, s' \in R$ and assume that $r + I = r' + I$ and $s + I = s' + I$. Then

$$rs - r's' = rs - r's + r's - r's' = (r - r')s + r'(s - s') \in I,$$

where we have used that $r - r' \in I$ and $s - s' \in I$ as well as the absorption property of the ideal I. Thus, $rs + I = r's' + I$ and the formula for the product of cosets is independent of chosen coset representatives, implying that it is well defined.

(ii) We know from group theory that $(R/I, +)$ is a group, and the formula for sum of cosets shows that this group is abelian. Let $r+I, s+I, t+I \in R/I$. Then

$$\begin{aligned}(r + I)[(s + I)(t + I)] &= (r + I)(st + I) = r(st) + I = (rs)t + I\\ &= (rs + I)(t + I) = [(r + I)(s + I)](t + I),\end{aligned}$$

so multiplication of cosets is associative. Similar arguments show that the two distributive laws hold. Therefore, $(R/I, +, \cdot\,)$ is a ring. □

$(R/I, +, \cdot\,)$ is the **quotient ring** (or **factor ring**) of R by I. The notation R/I is read "R modulo I" or "R mod I."

Recall that $0 + I$ $(= I)$ is the additive identity of R/I and the additive inverse of $r + I \in R/I$ is $-r + I$. If R has identity 1, then R/I has identity $1 + I$ provided $1 + I \neq 0 + I$, which is the case if and only if I is proper.

The assumption that I is an ideal (rather than simply a subring) is essential here. Indeed if I fails to have the absorption property, then the product given in the theorem is not well-defined (see Exercise 5–1).

5.2 Example: Z/nZ

Let n be a positive integer. Since $n\mathbf{Z}$ is an ideal of $\mathbf{Z}$ (in fact, $n\mathbf{Z} = (n)$), the quotient ring $\mathbf{Z}/n\mathbf{Z}$ is defined. The group isomorphism $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}_n$ given by $m + n\mathbf{Z} \mapsto r$, where r is the remainder of m upon division by n, is an isomorphism of rings as well. Therefore

$$\mathbf{Z}/n\mathbf{Z} \cong \mathbf{Z}_n.$$

For example, taking n to be 3 and simplifying notation by putting $I = 3\mathbf{Z}$, we have $\mathbf{Z}/3\mathbf{Z} = \mathbf{Z}/I = \{0+I, 1+I, 2+I\}$ and the operation tables for this ring are

$+$	$0+I$	$1+I$	$2+I$
$0+I$	$0+I$	$1+I$	$2+I$
$1+I$	$1+I$	$2+I$	$0+I$
$2+I$	$2+I$	$0+I$	$1+I$

$\cdot$	$0+I$	$1+I$	$2+I$
$0+I$	$0+I$	$0+I$	$0+I$
$1+I$	$0+I$	$1+I$	$2+I$
$2+I$	$0+I$	$2+I$	$1+I$

If every occurrence of $+I$ is suppressed, one obtains the operation tables of the ring $\mathbf{Z}_3 = \{0, 1, 2\}$, which demonstrates that $\mathbf{Z}/3\mathbf{Z} \cong \mathbf{Z}_3$ as expected.

5.3 Theorems of Euler and Fermat

Let n be a positive integer. For $a, b \in \mathbf{Z}$, we say that a is **congruent modulo** n to b, written $a \equiv b \pmod{n}$, if the difference $a - b$ is divisible by n.

For example, $2 \equiv 12 \pmod 5$, since $2 - 12 = -10$ is divisible by 5.

Congruence modulo n is an equivalence relation on $\mathbf{Z}$. In fact, it is the same relation on $\mathbf{Z}$ as (right) congruence modulo the subgroup $n\mathbf{Z}$ (defined by putting $a \equiv_r b \pmod{n\mathbf{Z}}$ if and only if $a - b \in n\mathbf{Z}$).

In the following theorem, ϕ is the Euler phi function (Section 2.5).

5.3.1 Theorem.

(i) (Euler's theorem) *For every integer m with $\gcd(m, n) = 1$, we have $m^{\phi(n)} \equiv 1 \pmod n$.*

(ii) (Fermat's little theorem) *For every integer m and prime number p, we have $m^p \equiv m \pmod p$.*

Proof. (i) Let m be an integer with $\gcd(m, n) = 1$. Let r be the remainder of m upon division by n. Then $m = qn + r$ for some integer q, so any divisor of both n and r is also a divisor of m. Since $\gcd(m, n) = 1$, it follows that $\gcd(r, n) = 1$. Since also $0 \leq r < n$, it follows from Section 2.5 that r is in the group of units of $\mathbf{Z}_n$, which has order $\phi(n)$. A corollary of Lagrange's theorem says that any element of a finite group raised to the order of the group is equal to the identity. Thus, $r^{\phi(n)} = 1$. Now $m + n\mathbf{Z}$ corresponds to r under the isomorphism $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}_n$ described in Section 5.2, so $m^{\phi(n)} + n\mathbf{Z} = (m + n\mathbf{Z})^{\phi(n)} = 1 + n\mathbf{Z}$. Therefore, $m^{\phi(n)} - 1 \in n\mathbf{Z}$, which implies that $m^{\phi(n)} \equiv 1 \pmod n$.

(ii) Let m be an integer and let p be a prime number. If $p \nmid m$ then (i) applies with $n = p$ to give $m^{p-1} \equiv 1 \pmod p$ and multiplying both sides by m gives the stated equivalence. If $p \mid m$, then both sides of the stated equivalence are congruent to 0 modulo p. □

5.4 Ideal is prime iff quotient is integral domain

Let R be a commutative ring with identity. An ideal I of R is a **prime ideal** if

(i) $I \neq R$,

(ii) $rs \in I \ (r, s \in R) \implies r \in I$ or $s \in I$.

In other words, an ideal I is prime provided it is proper and the only way a product can be in I is if one of the factors is in I.

Let I be an ideal of R.

5.4.1 Theorem. *The ideal I is prime if and only if R/I is an integral domain.*

Proof. ($\Rightarrow$) Assume that I is prime. The ring R/I is commutative and it has identity $1 + I$ (for this we need to observe that $1 + I \neq 0 + I$ since I is proper). Let $r + I, s + I \in R/I$ and assume that $(r + I)(s + I) = 0 + I$ and $r + I \neq 0 + I$. Now $rs + I = (r + I)(s + I) = I$, implying $rs \in I$. Since I is prime, and $r \notin I$, we have $s \in I$, so that $s + I = 0 + I$. Thus R/I is an integral domain.

($\Leftarrow$) Assume that R/I is an integral domain. Then R/I has an identity different from the zero element $0 + I$. In particular, R/I has at least two elements, implying that $I \neq R$. Let $r, s \in R$ and assume that $rs \in I$ and $r \notin I$. Then $(r + I)(s + I) = rs + I = 0 + I$ and $r + I \neq 0 + I$. Since R/I is an integral domain, we conclude that $s + I = 0 + I$ so that $s \in I$. Therefore, I is prime. □

Let n be a positive integer. Since $\mathbf{Z}/(n) = \mathbf{Z}/n\mathbf{Z} \cong \mathbf{Z}_n$ is an integral domain if and only if n is prime (see Section 4.4), it follows from the theorem that (n) is a prime ideal of $\mathbf{Z}$ if and only if n is prime.

The notion of a prime ideal provides another characterization of integral domain, namely, R is an integral domain if and only if the ideal $\{0\}$ is prime. This statement follows directly from the definitions, but it can also be seen using the theorem since $R/\{0\} \cong R$.

5.5 Ideal is maximal iff quotient is field

Let R be a commutative ring with identity. An ideal I of R is a **maximal ideal** if

(i) $I \neq R$,

(ii) $I \subsetneqq J \lhd R \implies J = R$.

In other words, an ideal I is maximal if and only if it is proper and it is not properly contained in any other proper ideal of R.

Let I be an ideal of R.

5.5.1 Theorem. *The ideal I is maximal if and only if R/I is a field.*

Proof. ($\Rightarrow$) Assume that I is maximal. Since I is proper, we have $1+I \neq 0+I$, so R/I has identity $1+I$. Also, R/I is commutative since R is commutative. Let $s+I$ be a nonzero element of R/I. Put

$$J = (s) + I = Rs + I = \{rs + a \,|\, r \in R, a \in I\},$$

which is an ideal by Exercise 3–3. For $a \in I$ we have $a = 0s + a \in J$, so $I \subseteq J$. Now $s = 1s + 0 \in J$, but $s \notin I$ (since $s + I \neq I$), so $I \subsetneqq J \lhd R$. Since I is maximal, it follows that $J = R$. Thus $1 \in R = J$, implying that $1 = rs + a$ for some $r \in R$ and $a \in I$. Then $(r+I)(s+I) = rs + I = 1 + I$ (since $1 - rs = a \in I$), so $s + I$ is a unit.

($\Leftarrow$) Assume that R/I is a field. Then R/I has an identity different from the zero element $0+I$. In particular, R/I has at least two elements, implying that $I \neq R$. Let I and J be ideals of R and assume that $I \subsetneqq J \lhd R$. Since $I \subsetneqq J$, there exists $r \in J \backslash I$. Then $r + I \neq 0 + I$, so there exists $s + I \in R/I$ such that $1 + I = (r+I)(s+I) = rs + I$ (since every nonzero element of R/I is a unit). We have $1 - rs \in I \subseteq J$, so $1 = (1 - rs) + rs \in J$, implying that $J = R$ (see Section 3.3). Thus, I is maximal. □

Let n be a positive integer. Since $\mathbf{Z}/(n) = \mathbf{Z}/n\mathbf{Z} \cong \mathbf{Z}_n$ is a field if and only if n is prime (see Section 2.6), it follows from the theorem that (n) is a maximal ideal of $\mathbf{Z}$ if and only if n is prime. In view of Section 5.4, we see that the maximal ideals and the prime ideals of $\mathbf{Z}$ coincide.

We see in the next section that every maximal ideal is prime, but not conversely.

5.6 Maximal ideal is prime

Let R be a commutative ring with identity and let I be an ideal of R.

5.6.1 Theorem. *If I is maximal, then I is prime.*

Proof. Assume that I is maximal. Then R/I is a field (Section 5.5) and hence an integral domain (Section 4.3). Therefore, I is prime (Section 5.4). □

The converse of this theorem is not true. In other words, it is possible for an ideal to be prime but not maximal. For instance, considering the ideal $I = \{0\}$ of the ring $\mathbf{Z}$, we have $\mathbf{Z}/I \cong \mathbf{Z}$, which is an integral domain but not a field. Therefore, I is prime by 5.4.1, but it is not maximal by 5.5.1.

5 – Exercises

5–1 Let R be a ring and let S be a subring of R that is *not* an ideal. Prove that there are cosets $r + S$ and $t + S$ of S for which the product formula $(r + S)(t + S) = rt + S$ is not well defined.

5–2 Let n be a nonnegative integer and assume that $p = 4n + 3$ is prime. Use Fermat's little theorem (5.3) to show that there is no integer x for which $x^2 \equiv -1 \pmod{p}$.

5–3 Let R be a commutative ring with identity and let S be the set consisting of 0 and all of the divisors of zero in R. Prove that S contains at least one prime ideal of R.

Hint: Consider an ideal of R that is maximal among all ideals of R contained in S (one can show that such an ideal exists by Zorn's lemma, but you may just assume existence). Use Exercise 3–3.

5–4 Let R be a commutative ring with identity. Use Section 5.5 to prove that if $\{0\}$ and R are the only ideals of R, then R is a field. (Cf. Exercise 4–2.)

6 Homomorphism

6.1 Definitions

Let R and R' be rings. A **homomorphism** from R to R' is a function $\varphi : R \to R'$ satisfying

(i) $\varphi(r+s) = \varphi(r) + \varphi(s)$,

(ii) $\varphi(rs) = \varphi(r)\varphi(s)$,

for all $r, s \in R$.

- A **monomorphism** is an injective homomorphism.
- An **epimorphism** is a surjective homomorphism.
- An **isomorphism** is a bijective homomorphism.
- An **automorphism** is an isomorphism from a ring to itself.

"Isomorphism" was defined earlier in Section 2.9. Recall that two rings R and R' are **isomorphic**, written $R \cong R'$, if there exists an isomorphism from one to the other. In this case, the rings R and R' are identical as far as their ring properties are concerned.

6.2 Examples

- Let n be a positive integer. The function $\varphi : \mathbf{Z} \to \mathbf{Z}_n$ given by $\varphi(m) = r$, where r is the remainder of m upon division by n, is a homomorphism, the **reduction modulo n homomorphism**.
- Let R be a commutative ring with identity and let $r \in R$. For a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i \in R[x]$, denote by $f(r)$ the element of R obtained by replacing x with r (so $f(r) = \sum_{i=0}^{n} a_i r^i$). The function $\varphi_r : R[x] \to R$ given by $\varphi_r(f(x)) = f(r)$ is a homomorphism, the **evaluation homomorphism** determined by r.
- Let R and R' be rings with identity and let $\sigma : R \to R'$ be a ring homomorphism. For a polynomial $f(x) \in R[x]$, denote by $\sigma f(x)$ the

polynomial in $R'[x]$ obtained by applying σ to the coefficients of $f(x)$. In symbols, if $f(x) = \sum_i a_i x^i$, then $\sigma f(x) = \sum_i \sigma(a_i)x^i$. The function $\bar{\sigma} : R[x] \to R'[x]$ given by $\bar{\sigma}(f(x)) = \sigma f(x)$ is a homomorphism, the **homomorphism induced by** σ. (See Section 9.8.)

- Let R be a ring and let I be an ideal of R. The function $\pi : R \to R/I$ given by $\pi(r) = r + I$ is an epimorphism, the **canonical epimorphism**.

6.3 Elementary properties

Recorded here are some standard facts about ring homomorphisms. The statements are either reiterations of ones already known from group theory (a ring homomorphism is a group homomorphism after all), or ring-theoretical analogs of such.

Let $\varphi : R \to R'$ be a ring homomorphism.

6.3.1 Theorem.

(i) $\varphi(0) = 0$.

(ii) $\varphi(-r) = -\varphi(r)$ *for each* $r \in R$.

(iii) *If* $S \leq R$, *then* $\varphi(S) \leq R'$.

(iv) *If* $S' \leq R'$, *then* $\varphi^{-1}(S') \leq R$.

Proof. Parts (i) and (ii) are known from group theory.

(iii) Let $S \leq R$. The set $\varphi(S) = \{\varphi(s) \,|\, s \in S\}$ is known to be a subgroup of $(R', +)$ by group theory. For every $s, t \in S$, we have $st \in S$, so $\varphi(s)\varphi(t) = \varphi(st) \in \varphi(S)$. This shows that $\varphi(S)$ is closed under multiplication. Therefore, it is a subring of R'.

(iv) Let $S' \leq R'$. The set $\varphi^{-1}(S') = \{r \in R \,|\, \varphi(r) \in S'\}$ is known to be a subgroup of $(R, +)$ by group theory. For every $r, s \in \varphi^{-1}(S')$, we have $\varphi(r), \varphi(s) \in S'$, so $\varphi(rs) = \varphi(r)\varphi(s) \in S'$, implying $rs \in \varphi^{-1}(S')$. This shows that $\varphi^{-1}(S')$ is closed under multiplication. Therefore, it is a subring of R. □

6.4 Kernel and image

Let $\varphi : R \to R'$ be a homomorphism of rings. The kernel of φ and the image of φ retain their same meanings from group theory:

- $\ker \varphi = \varphi^{-1}(\{0\}) = \{a \in R \,|\, \varphi(a) = 0\}$, the **kernel** of φ,
- $\operatorname{im} \varphi = \varphi(R) = \{\varphi(a) \,|\, a \in R\}$, the **image** of φ.

The kernel of φ is a subring of R by Theorem 6.3(iv) with $S' = \{0\}$. In fact, the kernel of φ is even an ideal of R as we will see in the next section. The image of φ is a subring of R' by Theorem 6.3(iii) with $S = R$.

6.5 Kernel same thing as ideal

Let R be a ring. The following theorem says that the notions "kernel of a homomorphism from R" and "ideal of R" amount to the same thing.

6.5.1 Theorem. *If $\varphi : R \to R'$ is a homomorphism, then $\ker \varphi$ is an ideal of R. Conversely, if I is an ideal of R, then I is the kernel of a homomorphism, namely, the canonical epimorphism $\pi : R \to R/I$.*

Proof. Let $\varphi : R \to R'$ be a homomorphism. It was observed in 6.4 that $\ker \varphi$ is a subring of R. If $a \in \ker \varphi$ and $r \in R$, then $\varphi(ra) = \varphi(r)\varphi(a) = \varphi(r) \cdot 0 = 0$, so $ra \in \ker \varphi$. This says that $\ker \varphi$ has the left absorption property. A similar proof shows that it has the right absorption property as well, so it is an ideal of R.

Let I be an ideal of R. An element r of R is in the kernel of π if and only if $I = \pi(r) = r + I$, which occurs if and only if $r \in I$. Therefore, $I = \ker \pi$. $\square$

6.6 Homomorphism is injective iff kernel is trivial

Here is a reminder of a useful criterion from group theory for checking injectivity of a homomorphism.

Let $\varphi : R \to R'$ be a homomorphism of rings.

6.6.1 Theorem. *The homomorphism φ is injective if and only if* $\ker \varphi = \{0\}$.

Proof. Assume that φ is injective. Let $r \in \ker \varphi$. Then $\varphi(r) = 0$. But also, $\varphi(0) = 0$ by Section 6.3. So $\varphi(r) = \varphi(0)$ and injectivity of φ gives $r = 0$. This shows that $\ker \varphi \subseteq \{0\}$. Since a kernel is a subgroup, the other inclusion is immediate.

Now assume that $\ker \varphi = \{0\}$. Let $r, s \in R$ and assume that $\varphi(r) = \varphi(s)$. Then $\varphi(r - s) = \varphi(r) - \varphi(s) = 0$, implying that $r - s \in \ker \varphi = \{0\}$. Thus, $r - s = 0$, that is, $r = s$. Therefore, φ is injective. □

As a practical matter, we observe that, in order to show that a homomorphism φ is injective, it suffices to show that $\ker \varphi \subseteq \{0\}$, since the other inclusion always holds ($\ker \varphi$ is a subgroup).

6.7 Fundamental Homomorphism Theorem

This theorem has an analog for groups. It is a generalization, useful in its own right, of the main step in the proof of the First Isomorphism Theorem.

Let $\varphi : R \to R'$ be a homomorphism of rings.

6.7.1 Theorem (Fundamental Homomorphism Theorem). *Let I be an ideal of R with $I \subseteq \ker \varphi$. There exists a unique homomorphism*

$$\overline{\varphi} : R/I \to R'$$

such that $\overline{\varphi}\pi = \varphi$, where $\pi : R \to R/I$ is the canonical epimorphism. The function $\overline{\varphi}$ is given by $\overline{\varphi}(r + I) = \varphi(r)$.

Proof. As in the statement of the theorem, let $\overline{\varphi} : R/I \to R'$ be the function given by $\overline{\varphi}(r + I) = \varphi(r)$.

If $r + I = s + I$ $(r, s \in R)$, then $r - s \in I \subseteq \ker \varphi$, so that $\varphi(r) - \varphi(s) = \varphi(r - s) = 0$, implying $\varphi(r) = \varphi(s)$. Thus, $\overline{\varphi}$ is well defined.

For $r + I, s + I \in R/I$, we have

$$\begin{aligned}\overline{\varphi}((r + I) + (s + I)) &= \overline{\varphi}((r + s) + I) = \varphi(r + s)\\ &= \varphi(r) + \varphi(s) = \overline{\varphi}(r + I) + \overline{\varphi}(s + I)\end{aligned}$$

and similarly $\overline{\varphi}((r+I)(s+I)) = \overline{\varphi}(r+I)\overline{\varphi}(s+I)$, so $\overline{\varphi}$ is a homomorphism.

For $r \in R$, we have

$$(\overline{\varphi}\pi)(r) = \overline{\varphi}(\pi(r)) = \overline{\varphi}(r + I) = \varphi(r),$$

giving $\overline{\varphi}\pi = \varphi$.

Finally, let $\psi : R/I \to R'$ be a homomorphism such that $\psi\pi = \varphi$. Then for any $r + I \in R/I$ we have

$$\psi(r + I) = \psi(\pi(r)) = (\psi\pi)(r) = \varphi(r) = \overline{\varphi}(r + I),$$

so that $\psi = \overline{\varphi}$, thus establishing uniqueness. □

6 – Exercises

6–1 Let R and R' be rings and let $\varphi : R \to R'$ be a homomorphism. Assume that R has identity 1.

(a) Prove that if φ is surjective, then $\varphi(1)$ is an identity element for R'.

(b) Give an example to show that if φ is not surjective, then $\varphi(1)$ need not be an identity element for R'.

6–2 Let R and R' be rings and let $\varphi : R \to R'$ be a homomorphism.

(a) Prove that if J is an ideal of R', then $\varphi^{-1}(J)$ is an ideal of R.

(b) Prove that if I is an ideal of R and φ is surjective, then $\varphi(I)$ is an ideal of R'.

(c) Give an example to show that, without the assumption of surjectivity in (b), $\varphi(I)$ need not be an ideal of R'.

6–3 Let R be a ring and let $R' = \mathrm{End}(R)$ be the endomorphism ring of the underlying abelian group of R (see 1.8). For $r \in R$, define a function $\lambda_r : R \to R$ by $\lambda_r(s) = rs$.

(a) Prove that $\lambda_r \in R'$ for each $r \in R$.

(b) Prove that the function $\varphi : R \to R'$ given by $\varphi(r) = \lambda_r$ is a homomorphism.

(c) Prove that the function φ is a monomorphism if R has an identity.

7 Isomorphism theorems

7.1 First Isomorphism Theorem

Let $\varphi : R \to R'$ be a homomorphism of rings. By Theorem 6.5, $\ker \varphi$ is an ideal of R so the quotient ring $R/\ker \varphi$ is defined.

7.1.1 Theorem (First Isomorphism Theorem).

$$R/\ker \varphi \cong \operatorname{im} \varphi.$$

Proof. Put $I = \ker \varphi$. By Section 6.7, the function $\overline{\varphi} : R/I \to R'$ given by $\overline{\varphi}(r + I) = \varphi(r)$ is a well-defined homomorphism. By restricting the codomain to $\operatorname{im} \varphi$ we obtain an epimorphism $R/I \to \operatorname{im} \varphi$, which we continue to denote by $\overline{\varphi}$.

Let $r + I \in \ker \overline{\varphi}$. Then $\varphi(r) = \overline{\varphi}(r + I) = 0$, so that $r \in \ker \varphi = I$. Thus, $r + I = I$. This shows that $\ker \overline{\varphi} \subseteq \{I\}$ so that $\overline{\varphi}$ is injective (see 6.6).

Therefore, $\overline{\varphi} : R/I \to \operatorname{im} \varphi$ is an isomorphism and $R/\ker \varphi = R/I \cong \operatorname{im} \varphi$. □

7.2 Quotient same thing as homomorphic image

Let R be a ring. The following theorem says that the notions "quotient of R" and "homomorphic image of R" amount to the same thing.

7.2.1 Theorem. *If R/I ($I \lhd R$) is a quotient of R, then R/I is a homomorphic image of R, namely, the image under the canonical epimorphism $\pi : R \to R/I$. Conversely, the image $\operatorname{im} \varphi = \varphi(R)$ of R under a homomorphism $\varphi : R \to R'$ is isomorphic to a quotient of R, namely $R/\ker \varphi$.*

Proof. Since the canonical epimorphism $\pi : R \to R/I$ is surjective, its image is R/I, so the first statement follows. The second statement is given by the First Isomorphism Theorem (7.1). □

7.3 Second Isomorphism Theorem

Let R be a ring, let S be a subring of R and let I be an ideal of R. By Exercise 7–1, $S + I$ is a subring of R. It contains I as an ideal so the quotient ring $(S+I)/I$ is defined. Also, $S \cap I$ is an ideal of S (as is easily checked), so the quotient ring $S/(S \cap I)$ is defined.

7.3.1 Theorem (Second Isomorphism Theorem).

$$S/(S \cap I) \cong (S+I)/I.$$

Proof. Define $\varphi : S \to (S+I)/I$ by $\varphi(s) = s+I$. Then φ is a homomorphism (it is simply the restriction to S of the canonical epimorphism $R \to R/I$).

For $s \in S$ we have

$$s \in \ker \varphi \iff \varphi(s) = I \iff s + I = I \iff s \in S \cap I,$$

so $\ker \varphi = S \cap I$.

Let $x \in (S+I)/I$. Then $x = (s+a) + I = s + I$ for some $s \in S$ and $a \in I$, and we have $\varphi(s) = s + I = x$, so φ is surjective.

By the First Isomorphism Theorem (7.1),

$$S/(S \cap I) = S/\ker \varphi \cong \operatorname{im} \varphi = (S+I)/I,$$

and the proof is complete. □

7.4 Third Isomorphism Theorem

Let R be a ring and let I and J be ideals of R with $J \supseteq I$. Then J/I is an ideal of R/I (as is easily checked), so the quotient ring $(R/I)/(J/I)$ is defined.

7.4.1 Theorem (Third Isomorphism Theorem).

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

Proof. Let $\psi : R \to R/J$ be the canonical epimorphism. Since $I \subseteq J = \ker \psi$, the Fundamental Homomorphism Theorem (6.7) says that the function $\varphi : R/I \to R/J$ given by $\varphi(r+I) = \psi(r) = r + J$ is a well-defined homomorphism. It follows from the indicated formula that φ is surjective.

We claim that $\ker \varphi = J/I$. Let $r + I \in R/I$. We first observe that if $r + I \in J/I$, then $r + I = b + I$ for some $b \in J$, implying $r = b + a \in J$ for some $a \in I$. Using this observation to supply the direction $\Leftarrow$ of the final step, we have

$$\begin{aligned} r + I \in \ker\varphi &\iff \varphi(r+I) = J \iff r + J = J \\ &\iff r \in J \iff r + I \in J/I, \end{aligned}$$

so the claim is established.

By the First Isomorphism Theorem (7.1),

$$(R/I)/(J/I) = (R/I)/\ker\varphi \cong \operatorname{im}\varphi = R/J,$$

and the proof is complete. □

7.5 Correspondence Theorem

Let $\varphi : R \to R'$ be an epimorphism of rings. Let

$$\begin{aligned} \mathbf{A} &= \{S \mid \ker\varphi \subseteq S \leq R\}, \\ \mathbf{A}' &= \{S' \mid S' \leq R'\}. \end{aligned}$$

In the statement of the following theorem, the notation $|T : S|$ has the same meaning that it had in group theory; it is the cardinality of the set of (left) cosets of S in T.

7.5.1 Theorem (Correspondence Theorem).

(a) *The map $\mathbf{A} \to \mathbf{A}'$ given by $S \mapsto \varphi(S)$ is a well-defined bijection. Its inverse map $\mathbf{A}' \to \mathbf{A}$ is given by $S' \mapsto \varphi^{-1}(S')$.*

(b) *For $S, T \in \mathbf{A}$, $\varphi(S) \subseteq \varphi(T)$ if and only if $S \subseteq T$, and in this case $|\varphi(T) : \varphi(S)| = |T : S|$.*

(c) *For $S, T \in \mathbf{A}$, $\varphi(S) \triangleleft \varphi(T)$ if and only if $S \triangleleft T$, and in this case $\varphi(T)/\varphi(S) \cong T/S$.*

Proof. (a) By 6.3, if S is a subring of R, then $\varphi(S)$ is a subring of R' so the map is well defined. By this same section, if S' is a subring of R', then $\varphi^{-1}(S')$ is a subring of R, and this latter subring contains $\ker\varphi$ since $\varphi(k) = 0 \in S'$ for all $k \in \ker\varphi$. Therefore, the indicated inverse map is also well defined. It suffices to show that both compositions of these two functions yield the respective identity maps on $\mathbf{A}$ and $\mathbf{A}'$.

Let $S \in \mathbf{A}$. We need to show that $\varphi^{-1}(\varphi(S)) = S$. Let $r \in \varphi^{-1}(\varphi(S))$. Then $\varphi(r) \in \varphi(S)$, implying that $\varphi(r) = \varphi(s)$ for some $s \in S$. Therefore,

$$\varphi(r - s) = \varphi(r) - \varphi(s) = 0$$

so that $r - s \in \ker \varphi \subseteq S$. It follows that $r \in S$. This gives $\varphi^{-1}(\varphi(S)) \subseteq S$. The other inclusion is immediate.

Let $S' \in \mathbf{A}'$. We need to show that $\varphi(\varphi^{-1}(S')) = S'$. Let $s' \in S'$. Since φ is surjective, there exists $r \in R$ such that $\varphi(r) = s'$. But this last equation says that $r \in \varphi^{-1}(S')$, so $s' \in \varphi(\varphi^{-1}(S'))$. This gives $S' \subseteq \varphi(\varphi^{-1}(S'))$. The other inclusion is immediate.

(b) Let $S, T \in \mathbf{A}$. If $\varphi(S) \subseteq \varphi(T)$, then, using (a), we have

$$S = \varphi^{-1}(\varphi(S)) \subseteq \varphi^{-1}(\varphi(T)) = T,$$

and the other implication is immediate.

Assume that $S \subseteq T$. We claim that the map $f : \{t + S \mid t \in T\} \to \{\varphi(t) + \varphi(S) \mid t \in T\}$ given by $f(t+S) = \varphi(t) + \varphi(S)$ is a well-defined bijection. For $t, t' \in T$, we have

$$\begin{aligned} t + S = t' + S \Rightarrow t - t' \in S \Rightarrow \varphi(t) - \varphi(t') = \varphi(t - t') \in \varphi(S) \\ \Rightarrow \varphi(t) + \varphi(S) = \varphi(t') + \varphi(S), \end{aligned}$$

so f is well-defined. Let $t, t' \in T$ and suppose that $f(t+S) = f(t'+S)$. Then $\varphi(t) + \varphi(S) = \varphi(t') + \varphi(S)$, implying that $\varphi(t - t') = \varphi(t) - \varphi(t') \in \varphi(S)$. Therefore, $t - t' \in \varphi^{-1}(\varphi(S)) = S$, and so $t + S = t' + S$. This shows that f is injective. That f is surjective is immediate, so the claim that f is bijective is established. We conclude that the domain and the codomain of f have the same cardinality, that is, $|T : S| = |\varphi(T) : \varphi(S)|$.

(c) Let $S, T \in \mathbf{A}$ and assume that $\varphi(S) \triangleleft \varphi(T)$. Using part (a) and Exercise 6–2, we have

$$S = \varphi^{-1}(\varphi(S)) \triangleleft \varphi^{-1}(\varphi(T)) = T.$$

The other implication also follows from Exercise 6–2.

Assume that $S \triangleleft T$. The restriction of φ to T composed with the canonical epimorphism yields an epimorphism $T \to \varphi(T) \to \varphi(T)/\varphi(S)$ having kernel $\varphi^{-1}(\varphi(S)) = S$, so the First Isomorphism Theorem (7.1) completes the proof. □

7 – Exercises

7–1 Let R be a ring, let S be a subring of R and let I be an ideal of R. Prove that $S+I$ is a subring of R, where $S+I = \{s+a \mid s \in S \text{ and } a \in I\}$.

7–2 Let R be a commutative ring with identity. Prove that $R[x]/(x) \cong R$.

7–3 Let R be a ring and put $I = \{(r,0) \mid r \in R\}$. Prove that I is an ideal of $R \oplus R$ and $(R \oplus R)/I \cong R$.

8 Factorization in an integral domain

8.1 Motivation

8.1.1 Theorem (Fundamental Theorem of Arithmetic). *An integer greater than one can be factored as a product of prime numbers, and such a factorization is unique up to the order of the factors.*

(Here, the word "product" has the usual broad meaning, which includes the possibility of a single factor.)

This theorem has many uses in the study of the ring of integers. For instance, it makes possible the notions of the greatest common divisor and the least common multiple of a collection of integers greater than one. Because of its usefulness, we seek a generalization of this theorem to other rings. For convenience, we will restrict our search to integral domains.

The theorem can be thought of as saying that prime numbers are the building blocks for the integers greater than one. We need a generalization of these building blocks to our arbitrary integral domain.

The reader most likely learned that a prime number is an integer greater than one having as positive factors only one and itself. This definition yields 2, 3, 5, 7, 11, and so on.

The words "greater than" and "positive" here create obstacles for any generalization, since the axioms for an integral domain provide no notions of order or positivity. Nor does it seem likely that reasonable such notions could exist in the light of examples of integral domains such as $\mathbf{Z}_5 = \{0, 1, 2, 3, 4\}$: if you start with 4 and add 2, you get 1, so it is not clear whether $4 > 1$ or $4 < 1$ makes more sense; also, $2 = -3$, so it is not clear whether 2 should be considered positive or negative.

The first thing to do is to stop insisting that a prime number be positive. After all, the negatives of the primes are as much like building blocks as the primes are (e.g., $6 = (-2)(-3)$). Also, the negatives of the primes can be used to build negative integers (e.g., $-6 = (-2)(3)$) and not just the integers greater than one as in the theorem.

With this in mind, we amend the old definition and say that a prime integer is any integer having the property that in any factorization of it, one of the factors must be either 1 or -1. The key observation is that 1 and -1 are precisely the units in $\mathbf{Z}$, and the notion of unit is defined using only the axioms for an integral domain.

As our definition stands, it includes 0 and ± 1 as prime integers. There is good reason to amend the definition to exclude these. The number 0 can only be a building block for 0 and there can be no unique factorization of 0 (e.g., $0 = (0)(2)$, $0 = (0)(2)(3)$). If we include 1 as a building block, then again we cannot get a statement of uniqueness of factorization (e.g., $2 = (1)(2)$, $2 = (1)(1)(2)$) and the same goes for -1. Therefore, we exclude 0 and ± 1 by insisting that a prime integer be neither zero nor a unit.

Summarizing,

- a **prime integer** is an integer p, neither zero nor a unit, having the property that in any factorization $p = mn$ ($m, n \in \mathbf{Z}$), either m or n is a unit.

A prime number has a second important property, namely, if it divides a product of two integers, then it must divide one or the other. For instance, the prime number 5 divides 30 and no matter how we express 30 as a product of integers, 5 always divides one of the factors:

$$\begin{aligned}
&30 = (2)(15) \text{ and } 5 \mid 15,\\
&30 = (-10)(-3) \text{ and } 5 \mid -10,\\
&30 = (6)(5) \text{ and } 5 \mid 5,
\end{aligned}$$

and so on. The nonprime 6 does not have this property, since $6 \mid 30$, but in the factorization $30 = (3)(10)$, we have $6 \nmid 3$ and $6 \nmid 10$.

So here is a second characterization:

- a **prime integer** is an integer p, neither zero nor a unit, having the property that if $p \mid mn$ $(m, n \in \mathbf{Z})$, then either $p \mid m$ or $p \mid n$.

These two characterizations of prime integer are equivalent and they describe the usual prime numbers together with their negatives: ± 2, ± 3, ± 5, ± 7, ± 11, and so on.

We alert the reader to a point of terminology. The natural generalizations of these two characterizations to other integral domains are not always equivalent, so in general one cannot use the same term for both properties as we have done here with the word "prime". For this reason, the term "prime" will be reserved for an element satisfying this second property, but an element satisfying the first property will be called "irreducible". (This is standard, but unfortunate since the first property is the one most closely related to our usual understanding of prime.)

The general definitions of "irreducible element" and "prime element" are given in Section 8.3. In Section 8.4 a definition is given for an integral domain in which a generalization of the Fundamental Theorem of Arithmetic holds (called a "unique factorization domain").

8.2 Divisibility and principal ideals

Let R be an integral domain and let r and s in R. We say that r **divides** s, written $r \mid s$, if $s = ra$ for some $a \in R$. We say that r and s are **associates**, written $r \sim s$, if $r = su$ for some unit $u \in R$. As the notation suggests, the property of being associates is an equivalence relation on R.

Here are some examples:

- Let $r \in R$. Then $r \mid 0$. However, $0 \mid r$ if and only if $r = 0$.

- $(R = \mathbf{Z}.)$ We have $2 \mid 6$ since $6 = (2)(3)$. For any integer n, we have $-n \sim n$, since $-n = n(-1)$ and -1 is a unit. In fact, for $n, m \in \mathbf{Z}$, we have $m \sim n$ if and only if $m = \pm n$.

- $(R = \mathbf{Q}.)$ Let $r, s \in \mathbf{Q}$. If $r \neq 0$, then $r \mid s$, since $s = r(s/r)$. If r and s are both nonzero, then $r \sim s$, since $r = s(r/s)$ and r/s is a unit.

- ($R = \mathbf{R}[x]$.) We have $(x+3) \mid (x^2+5x+6)$, since $x^2+5x+6 = (x+3)(x+2)$. Since the units in $\mathbf{R}[x]$ are precisely the nonzero constant polynomials, we have $f(x) \sim g(x)$ if and only if $f(x) = r \cdot g(x)$ for some nonzero $r \in \mathbf{R}$ ($f(x), g(x) \in \mathbf{R}[x]$).

We collect some elementary facts relating these new notions to principal ideals.

8.2.1 Theorem. *Let $r, s \in R$.*

(i) $r \mid s \iff s \in (r) \iff (s) \subseteq (r)$,

(ii) $r \sim s \iff r \mid s$ *and* $s \mid r$,

(iii) $(r) = (s) \iff r \sim s$,

(iv) $(r) = R \iff r$ *is a unit.*

Proof. (i) By Section 3.5 we have $(r) = Rr = rR$, so

$$r \mid s \iff s = ra,\ \text{some } a \in R \iff s \in (r) \iff (s) \subseteq (r).$$

(ii) Assume that $r \sim s$. Then $r = su$ for some unit $u \in R$. In particular, $s \mid r$. But also, $s = ru^{-1}$, so $r \mid s$ as well. Now assume that $r \mid s$ and $s \mid r$. Then $s = ra$ and $r = sb$ for some $a, b \in R$. This gives $r \cdot 1 = r = sb = (ra)b = r(ab)$. If $r \neq 0$, then cancellation (which is valid, since R is an integral domain) gives $1 = ab$, so that b is a unit and $r \sim s$. On the other hand, if $r = 0$, then $s = 0$ as well, so $r = s \cdot 1$ implying that $r \sim s$.

(iii) Using (i), we have $(r) = (s)$ if and only if $r \mid s$ and $s \mid r$, which holds if and only if $r \sim s$ by (ii).

(iv) Since $R = (1)$, we have $(r) = R$ if and only if $r \sim 1$ by (iii). Now $r \sim 1$ if and only if $r = 1u$ for some unit $u \in R$, which is the case if and only if r is a unit. □

8.3 Irreducible element, prime element

Let R be an integral domain and let r be a nonzero element of R. A unit u in R always allows for a trivial factorization of r, namely, $r = u(u^{-1}r)$. A factorization of r that is not of this trivial type is a **proper factorization**, that is, $r = st$ $(s, t \in R)$ is a proper factorization if neither s nor t is a unit. In the following definitions, we say that r is a nonzero nonunit to mean that r is neither zero nor a unit.

An element r of R is **irreducible** if

- r is a nonzero nonunit,
- r has no proper factorization.

An element r of R is **prime** if

- r is a nonzero nonunit,
- $r \mid st \quad (s, t \in R) \quad \Rightarrow \quad r \mid s \text{ or } r \mid t.$

8.3.1 Theorem. *Let $r \in R$.*

(i) *The element r is irreducible if and only if (r) is nonzero and maximal among the proper principal ideals of R.*

(ii) *The element r is prime if and only if the ideal (r) is nonzero and prime.*

(iii) *Every prime element of R is irreducible.*

(iv) *If every ideal of R is principal, then every irreducible element of R is prime.*

Proof. We make repeated use (without further reference) of Section 8.2.

(i) Assume that r is irreducible. First, (r) is nonzero, since r is nonzero, and it is proper since r is not a unit. Let s be an element of R such that $(r) \subseteq (s)$ and (s) is proper. Then $s \mid r$ so that $r = st$ for some $t \in R$. Now s is not a unit, since (s) is proper. Therefore t is a unit, since r is irreducible. Thus, $r \sim s$, implying that $(r) = (s)$. We conclude that (r) is nonzero and maximal among the proper principal ideals of R.

Now assume that (r) is nonzero and maximal among the proper principal ideals of R. Since (r) is nonzero and proper, r is a nonzero nonunit. Let $r = st$ $(s, t \in R)$ be a factorization of r and assume that s is a nonunit.

Then (s) is proper and $r \in (s)$, which implies that $(r) \subseteq (s)$. By maximality of (r), we have $(r) = (s)$, so that $r \sim s$. This implies that $r = su$ for some unit $u \in R$. Therefore, $st = su$ and by cancellation, $t = u$, so that t is a unit.

(ii) Assume that r is prime. Arguing as above, the ideal (r) is nonzero and proper. Let $s, t \in R$. Assume that $st \in (r)$ and $s \notin (r)$. Then $r \mid st$ and $r \nmid s$, implying that $r \mid t$ so that $t \in (r)$. This shows that the ideal (r) is prime.

Now assume that the ideal (r) is prime. Then, as above, r is a nonzero nonunit. Let $s, t \in R$. Assume that $r \mid st$ and $r \nmid s$. Then $st \in (r)$ and $s \notin (r)$, implying that $t \in (r)$ so that $r \mid t$. This shows that r is prime.

(iii) Let r be a prime element of R. First, r is a nonzero nonunit. Let $r = st$ $(s, t \in R)$ be a factorization of r. We have $st = r \cdot 1$, so $r \mid st$. Since r is prime, it follows that either $r \mid s$ or $r \mid t$. Suppose that $r \mid s$. Since also $s \mid r$, we get $r \sim s$ so that $r = su$ for some unit u of R. Then $st = su$, so that $t = u$ and t is a unit. The case $r \mid t$ similarly leads to the conclusion that s is a unit. Hence, r has no proper factorization, and it is therefore irreducible.

(iv) Assume that every ideal of R is principal and let r be an irreducible element of R. By part (i), (r) is maximal among the proper principal ideals of R, and since every ideal of R is principal, this says that (r) is a maximal ideal of R. By 5.6, (r) is prime, and so r is prime by part (ii). □

Since every ideal of $\mathbf{Z}$ is principal, the notions of irreducible integer and prime integer coincide and describe what we are used to calling the prime numbers along with their negatives, namely, ±2, ±3, ±5, ±7, ±11, and so on.

Although a prime element is always irreducible, it is not the case that every irreducible element is prime (see Exercise 8–1).

8.4 Unique factorization domain

Let R be an integral domain. We say that R is a **unique factorization domain**, or **UFD** for short, if the two following properties are satisfied:

(i) (Existence) Each nonzero nonunit of R can be written as a product of irreducible elements;

(ii) (Uniqueness) If $r \in R$ has two factorizations, $r = s_1 s_2 \cdots s_m$ and $r = t_1 t_2 \cdots t_n$ with each s_i and each t_i irreducible, then $m = n$ and for some permutation $\sigma \in \mathrm{Sym}(m)$ we have $s_i \sim t_{\sigma(i)}$ for each i.

(As always, "product" allows for the possibility of only one factor.) The uniqueness statement says that in any two factorizations of an element of R as a product of irreducible elements the factors can be reordered (if necessary) in such a way that corresponding factors are associates.

So, a UFD is an integral domain in which an analog of the Fundamental Theorem of Arithmetic holds. In particular, we expect the ring **Z** of integers to be a UFD. That this is the case is a consequence of the theorem of Section 8.5. Before stating it, though, we give an example to illustrate the two statements in the definition of UFD as applied to **Z**. We take r to be -120, a nonzero nonunit.

- The factorization $-120 = (-2)(2)(2)(3)(5)$, exhibits -120 as a product of irreducible elements.
- The factorization $-120 = (-3)(2)(5)(-2)(-2)$, also exhibits -120 as a product of irreducible elements. This second factorization can be reordered as $-120 = (-2)(-2)(2)(-3)(5)$ and comparing with the earlier factorization $-120 = (-2)(2)(2)(3)(5)$ we see that corresponding factors are indeed associates.

A field is a UFD (vacuously since a field has no nonzero nonunits). An example of an integral domain that is not a UFD is given in Exercises 8–1 and 8–4.

8.5 A PID is a UFD

An integral domain R is a **principal ideal domain**, or **PID** for short, if every ideal of R is principal.

8.5.1 Theorem. *Every PID is a UFD.*

Proof. Let R be a PID. We begin by proving the existence statement in 8.4 and the proof is by contradiction.

Suppose there is a nonzero nonunit r in R that cannot be written as a product of irreducible elements. Then r is not irreducible (since otherwise it is a product of a single irreducible element according to our convention). Therefore, r has a proper factorization $r = r_1 r_2$. Either r_1 or r_2 cannot be written as a product of irreducible elements and we are free to assume that this is the case for r_1. Section 8.2 shows that $(r) \subsetneqq (r_1)$. What we have shown is that a nonzero nonunit of R that cannot be written as product of irreducible elements generates an ideal that is properly contained in the ideal generated by another nonzero nonunit of R that cannot be written as a product of irreducible elements. We conclude that there exists a nonending sequence

$$(r) \subsetneqq (r_1) \subsetneqq (r_2) \subsetneqq \cdots$$

of principal ideals of R, each properly contained in the next. By Exercise 8–6, the union of these ideals is an ideal of R, and it is equal to (s) for some $s \in R$ since R is a PID. Now $s \in (r_i)$ for some i, so $(s) \subseteq (r_i)$. But then

$$(s) \subseteq (r_i) \subsetneqq (r_{i+1}) \subseteq (s),$$

a contradiction. This completes the proof of the existence statement.

Now we prove the uniqueness statement in 8.4. Let $r \in R$. Suppose that r has two factorizations, $r = s_1 s_2 \cdots s_m$ and $r = t_1 t_2 \cdots t_n$ with each s_i and each t_i irreducible. We proceed by induction on m, assuming, without loss of generality, that $m \geq n$. If $m = 1$, then $s_1 = r = t_1$ and the statement holds.

Assume that $m > 1$. We have

$$s_1 s_2 \cdots s_m = t_1 t_2 \cdots t_n,$$

so $s_m \mid t_1 t_2 \cdots t_n$. Since s_m is irreducible and R is a PID, s_m is prime (see 8.3). Therefore, $s_m \mid t_j$ for some j. By interchanging the factors t_n and t_j, if necessary, we may (and do) assume that $s_m \mid t_n$. Since t_n has no proper factorization, it must be the case that t_n and s_m are associates so that $s_m = u t_n$ for some unit $u \in R$. Substituting this equation into the earlier one and cancelling t_n from both sides yields

$$s_1 s_2 \cdots s_{m-2} s'_{m-1} = t_1 t_2 \cdots t_{n-1},$$

where $s'_{m-1} = s_{m-1} u$, which is irreducible by Exercise 8–2(a). By the induction hypothesis, $m - 1 = n - 1$ (implying that $m = n$) and there exists $\sigma \in \mathrm{Sym}(m-1)$ such that $s_i \sim t_{\sigma(i)}$ for all $1 \leq i < m$ (using that

$s_{m-1} \sim s'_{m-1}$). Viewing σ as an element of $\mathrm{Sym}(m)$ by putting $\sigma(m) = m$, we have $s_m \sim t_n = t_m = t_{\sigma(m)}$ as well, and the proof is complete. □

Since $\mathbf{Z}$ is a PID, the theorem says that $\mathbf{Z}$ is also a UFD. The Fundamental Theorem of Arithmetic now follows as a corollary: Let r be an integer greater than one. Then r is a nonzero nonunit, so the first property of UFD gives a factorization of r as a product of irreducibles, that is, a product of integers that are either prime numbers or negatives of prime numbers. Applying the absolute value to both sides and using the fact that $|mn| = |m||n|$ produces a factorization of r as a product of prime numbers. As for uniqueness, suppose we have two factorizations of r as a product of prime numbers. Since these prime numbers are irreducible, the second property of UFD insures that both factorizations have the same number of factors and that after rearranging the factors (if necessary) the corresponding factors are associates. Since two positive integers are associates only if they are equal, the uniqueness statement of the Fundamental Theorem of Arithmetic follows.

Every PID is a UFD, but it is not the case that every UFD is a PID. Indeed $\mathbf{Z}[x]$ is a UFD as we will see in Section 9.7, but it is not a PID (see Exercise 8–5).

8 – Exercises

8–1 Let $R = \{m + n\sqrt{10} \mid m, n \in \mathbf{Z}\}$ and define $N : R \to \mathbf{Z}$ by $N(m + n\sqrt{10}) = m^2 - 10n^2$.

(a) Prove that R is a subring of the ring $\mathbf{R}$ of real numbers.

(b) Prove that $N(rs) = N(r)N(s)$ for all $r, s \in R$.

(c) Prove that $r \in R$ is a unit if and only if $N(r) = \pm 1$.

(d) Prove that 2 is irreducible in R. (Hint: Assume 2 has a proper factorization and use the earlier parts and reduction modulo 5 to get a contradiction.)

(e) Prove that 2 is not prime in R. (Hint: $6 = (4 + \sqrt{10})(4 - \sqrt{10})$.)

8–2 Let R be an integral domain, let $r, s \in R$, and assume that $r \sim s$.

(a) Prove that if r is irreducible, then so is s.

(b) Prove that if r is prime, then so is s.

Hint: Instead of arguing from the definitions, consider using theorems.

8–3 Let R be a UFD. Prove that every irreducible element of R is prime.

8–4 Give two proofs that the integral domain R of Exercise 8–1 is not a UFD.

Hint: For an indirect proof, use 8–3. For a direct proof, use that $6 = (2)(3)$.

8–5 Prove that $\mathbf{Z}[x]$ is not a PID.

Hint: Suppose that $\mathbf{Z}[x]$ is a PID. Prove that the set I of all polynomials in $\mathbf{Z}[x]$ having even constant term is an ideal. By assumption, $I = (f(x))$ for some $f(x) \in \mathbf{Z}[x]$. Use the fact that $2 \in I$ to draw a conclusion about $f(x)$, and then use the fact that $x \in I$ to get a contradiction.

8–6 Let R be a ring and let $I_1, I_2, \ldots$ be a sequence of ideals of R with $I_i \subseteq I_{i+1}$ for each i. Prove that the union $I = \bigcup_i I_i$ is an ideal of R.

9 Polynomial ring

9.1 Definition

Let R be a ring with identity. The polynomial ring over R in the indeterminant x, denoted $R[x]$, was defined informally in Section 1.6. Here, we give a more rigorous definition.

Let $R[x]$ denote the set of all sequences $(a_i) = (a_0, a_1, a_2, \ldots)$ $(a_i \in R)$ that are eventually zero (meaning, there exists $m \in \mathbf{N}$ such that $a_i = 0$ for all

$i > m$). Define an addition and multiplication on $R[x]$ by

$$(a_i) + (b_i) = (a_i + b_i),$$
$$(a_i)(b_i) = (c_i), \quad \text{where} \quad c_i = \sum_{j=0}^{i} a_j b_{i-j}.$$

Then $(R[x], +, \cdot\,)$ is a ring.

Define x to be the sequence $(0, 1, 0, 0, \ldots)$. A simple proof by induction shows that for each positive integer i, x^i is the sequence with 1 in the ith position (with position labels starting at 0) and 0's elsewhere:

$$x^i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots).$$

Identifying $a \in R$ with the sequence $(a, 0, 0, \ldots)$ we have

$$ax^i = (a, 0, 0, \ldots)(0, \ldots, 0, \underset{i}{1}, 0, \ldots) = (0, \ldots, 0, \underset{i}{a}, 0, \ldots).$$

Therefore, if (a_i) is an element of $R[x]$, then there exists $m \in \mathbf{N}$ such that $a_i = 0$ for all $i > m$ and

$$\begin{aligned}(a_i) &= (a_0, a_1, a_2, \ldots, a_m, 0, \ldots)\\ &= \sum_{i=0}^{m} (0, \ldots, 0, \underset{i}{a_i}, 0, \ldots)\\ &= a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m.\end{aligned}$$

This brings us back to the traditional notation for polynomials as linear combinations of powers of x. Since such a linear combination is no longer just an "expression", which is a vague notion, but rather a sequence, the definition of a polynomial is now on firm ground.

It is traditional to use the linear combination form of a polynomial instead of the sequence form because of the notational and computational convenience the former affords. For instance, it is easier to write $7x^5$ than to write $(0, 0, 0, 0, 0, 7, 0, \ldots)$. Also, multiplying small polynomials in linear combination form by using the distributive law and collecting like terms is usually easier than applying the formula for multiplying sequences.

The reader may wonder why no fuss is made about the definition of a polynomial in a high school algebra course. It is not just that the level of sophistication required to understand a rigorous definition is deemed too high.

Rather, the sort of definition just given is not appropriate. In high school algebra, the polynomial $f(x) = \sum_{i=0}^{n} a_i x^i$ is really a function $f : \mathbf{R} \to \mathbf{R}$, which is not a vague concept.

If we attempted to define the polynomial $f(x) = \sum_{i=0}^{n} a_i x^i \in R[x]$ as the function $R \to R$ sending x to $\sum_{i=0}^{n} a_i x^i$ we would lose something. For instance, if $R = \mathbf{Z}_2$, then the polynomials x and x^2 are not equal (since their corresponding coefficients are not equal), but as functions $\mathbf{Z}_2 \to \mathbf{Z}_2$ they *are* equal since they both send $0 \mapsto 0$ and $1 \mapsto 1$. (For more on the process of passing from polynomials to polynomial functions, see Exercise 9–2.)

(Incidentally, this sort of collapsing does not happen for polynomial functions $\mathbf{R} \to \mathbf{R}$; due to properties of real numbers, the power functions $x \mapsto x^i$ ($i \in \mathbf{N} \cup \{0\}$) are linearly independent over $\mathbf{R}$, so two polynomial functions are equal if and only if their corresponding coefficients are equal.)

9.2 Degree of polynomial

Let R be an integral domain. Let $f(x)$ be a nonzero polynomial over R and write $f(x) = \sum_{i=0}^{n} a_i x^i$ with $a_n \neq 0$. The **degree** of $f(x)$, denoted $\deg f(x)$, is the integer n. In other words, the degree of a nonzero polynomial is the exponent of the highest power of x appearing in the polynomial (assuming that terms with zero coefficients are suppressed). For example, $\deg(2+4x^2-5x^6) = 6$.

It is convenient (as seen in the following theorem) to define the degree of the zero polynomial to be $-\infty$ and to use the natural conventions that $-\infty \leq n$ and $-\infty + n = -\infty$ for every integer n.

Let $f(x)$ and $g(x)$ be polynomials over R.

9.2.1 Theorem.

(i) $\deg[f(x) + g(x)] \leq \max\{\deg f(x), \deg g(x)\}$,

(ii) $\deg[f(x)g(x)] = \deg f(x) + \deg g(x)$.

Proof. (i) This follows from the observation that if neither $f(x)$ nor $g(x)$ has a term involving a particular power of x, then neither does their sum.

(ii) Let $m = \deg f(x)$ and $n = \deg g(x)$ and write $f(x) = \sum_{i=0}^{m} a_i x^i$ and $g(x) = \sum_{i=0}^{n} b_i x^i$. The term with the highest power of x appearing in the product $f(x)g(x)$ is $a_m b_n x^{m+n}$, and $a_m b_n \neq 0$ since $a_m, b_n \neq 0$ and R has no divisors of zero. Therefore, $\deg[f(x)g(x)] = m+n = \deg f(x) + \deg g(x)$.

□

The assumption in force is that R is an integral domain. If this is changed so that R is allowed to have divisors of zero, then (ii) is no longer valid. For instance, if $f(x) = 2x$ and $g(x) = 3x$, both polynomials over $\mathbf{Z}_6$, then $f(x)g(x) = 0$, whence

$$\deg[f(x)g(x)] = -\infty < 2 = \deg f(x) + \deg g(x).$$

9.3 Division algorithm

Consider the process of dividing 80 by 3 using long division. The algorithm is repeated until the remainder is less than the number being divided by. Here, we get an answer of 26 with a remainder of 2. Thus,

$$\frac{80}{3} = 26 + \frac{2}{3},$$

which can be written

$$80 = (26)(3) + 2.$$

This illustrates the following fact about the ring of integers:

Let m and n be integers with $n > 0$. There exist unique integers q and r with $0 \leq r < n$ such that $m = qn + r$.

In more familiar terms, m divided by n yields the quotient q with a remainder of r. This fact is known as the Division Algorithm for integers. (Actually, it is not an algorithm but rather a theorem; the long division steps for finding the q and r comprise the division algorithm.)

The reader is probably aware that there is also a long division algorithm for polynomials over $\mathbf{R}$. The algorithm is repeated until the remainder has degree less than the degree of the polynomial being divided by. This algorithm can be carried out for polynomials over any field:

9.3.1 Theorem (Division algorithm). *Let F be a field. If $f(x)$ and $g(x)$ are polynomials over F with $g(x) \neq 0$, then there exist unique polynomials $q(x)$ and $r(x)$ over F with $\deg r(x) < \deg g(x)$ such that*

$$f(x) = q(x)g(x) + r(x).$$

Proof. Let $f(x)$ and $g(x)$ be polynomials over F with $g(x) \neq 0$. We begin with the existence part of the proof. The set

$$S = \{f(x) - q(x)g(x) \mid q(x) \in F[x]\}$$

is nonempty (since $f(x) \in S$), so it has an element $r(x)$ of minimal degree (with $\deg r(x) = -\infty$ being a possibility). Now $r(x) = f(x) - q(x)g(x)$ for some $q(x) \in F[x]$, so $f(x) = q(x)g(x) + r(x)$ and we are done if we can prove that $\deg r(x) < \deg g(x)$.

Suppose, to the contrary, that $\deg r(x) \geq \deg g(x)$. We have

$$g(x) = \sum_{i=0}^{n} a_i x^i \quad \text{and} \quad r(x) = \sum_{i=0}^{m} b_i x^i$$

for some $a_i, b_i \in F$ with $a_n, b_m \neq 0$, where $n = \deg g(x)$ and $m = \deg r(x)$ (the expressions making sense since $m \geq n \geq 0$ due to the fact that $g(x) \neq 0$).

Since $a_n \neq 0$, the element a_n^{-1} is defined (F is a field), and since $m = \deg r(x) \geq \deg g(x) = n$, we have $x^{m-n} \in F[x]$. Therefore,

$$h(x) := r(x) - a_n^{-1} b_m x^{m-n} g(x)$$

is a polynomial over F. The formulas in Section 9.2 show that $\deg h(x) \leq m$ and, since the coefficient of x^m in $h(x)$ is $b_m - a_n^{-1} b_m a_n = 0$, we have $\deg h(x) < m = \deg r(x)$. Substituting $r(x) = f(x) - q(x)g(x)$ into the definition of $h(x)$ and regrouping gives

$$h(x) = f(x) - (q(x) + a_n^{-1} b_m x^{m-n}) g(x) \in S.$$

But this contradicts the definition of $r(x)$ as an element of S of minimal degree. We conclude that $\deg r(x) < \deg g(x)$ and the existence statement is established.

Next we turn to the uniqueness statement. Suppose that also $f(x) = q'(x)g(x) + r'(x)$ with $q'(x), r'(x) \in F[x]$ and $\deg r'(x) < \deg g(x)$. We have

$$r(x) - r'(x) = [f(x) - q(x)g(x)] - [f(x) - q'(x)g(x)] = [q'(x) - q(x)]g(x).$$

If $q(x) \neq q'(x)$, then we get the contradiction

$$\begin{aligned}\deg g(x) &> \max\{\deg r(x), \deg r'(x)\} \geq \deg[r(x) - r'(x)] \\ &= \deg[(q'(x) - q(x))g(x)] = \deg[q'(x) - q(x)] + \deg g(x) \\ &\geq \deg g(x),\end{aligned}$$

since $\deg[q'(x) - q(x)] \geq 0$. Therefore, $q'(x) = q(x)$ and the earlier equation gives $r'(x) = r(x)$ as well. This completes the proof. □

9.4 Zeros of a polynomial

Let R be a commutative ring with identity and let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial over R. An element r of R is a **zero** of $f(x)$ if $f(r) = 0$, where $f(r) = \sum_{i=0}^{n} a_i r^i$.

For example, the number 3 is a zero of the polynomial $f(x) = 6 - 5x + x^2 \in \mathbf{R}[x]$, since $f(3) = 6 - 5(3) + 3^2 = 0$.

9.4.1 Theorem. *Assume that R is a field.*

(i) *The element r of R is a zero of $f(x)$ if and only if $x - r$ divides $f(x)$.*

(ii) *If $f(x)$ is nonzero, then it has at most* $\deg f(x)$ *zeros in R.*

Proof. (i) Let $r \in R$. Assume that r is a zero of $f(x)$. By the division algorithm (9.3), there exist polynomials $q(x)$ and $s(x)$ over R with $\deg s(x) < \deg(x - r) = 1$ such that $f(x) = q(x)(x-r) + s(x)$. Now $\deg s(x)$ is 0 or $-\infty$, so in either case $s(x)$ is constant. Since $0 = f(r) = q(r)(r-r) + s(r) = s(r)$, it follows that $s(x)$ is the zero polynomial. Therefore, $f(x) = q(x)(x - r)$, and $x - r$ divides $f(x)$.

Now assume that $x - r$ divides $f(x)$. Then $f(x) = q(x)(x - r)$ for some $q(x) \in R[x]$. Thus, $f(r) = q(r)(r - r) = 0$, and r is a zero of $f(x)$.

(ii) Assume that $f(x)$ is nonzero, so in particular $\deg f(x) \geq 0$. We prove that $f(x)$ has at most $\deg f(x)$ zeros by using induction on the degree of $f(x)$. If $\deg f(x) = 0$, then $f(x)$ is a nonzero constant polynomial and therefore it has no zeros in accordance with the statement. Assume that $\deg f(x) > 0$. If $f(x)$ has no zeros, then the statement holds, so assume that $f(x)$ has a zero $r \in R$. By part (i), $x - r$ divides $f(x)$ so that $f(x) = q(x)(x - r)$ for some $q(x) \in R[x]$. If $s \in R$ is a zero of $f(x)$ and $s \neq r$, then s is a zero of $q(x)$ since $0 = f(s) = q(s)(s - r)$ and $s - r \neq 0$. It follows that $f(x)$ has at most one more zero than $q(x)$ has. Now $\deg q(x) = \deg f(x) - 1 < \deg f(x)$, so $q(x)$ has at most $\deg q(x)$ zeros by the induction hypothesis. Therefore, $f(x)$ has at most $\deg q(x) + 1 = \deg f(x)$ zeros, and the proof is complete. □

9.5 Irreducible polynomial

Let R be an integral domain and let $f(x) \in R[x]$. The polynomial $f(x)$ is **irreducible over** R if it is irreducible as an element of the integral domain $R[x]$, that is, it is a nonzero nonunit that does not have a proper factorization (see Section 8.3). A factorization is proper if neither factor is a unit, the units in this setting being precisely the units of R viewed as constant (degree zero) polynomials.

A polynomial that is a nonzero nonunit is **reducible** if it is not irreducible, that is, if it has a proper factorization.

- The polynomial $x^2 - 2$ is reducible over $\mathbf{R}$ since $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$, a proper factorization in $\mathbf{R}[x]$.
- The polynomial $x^2 + 1$ is reducible over $\mathbf{C}$ since $x^2 + 1 = (x + i)(x - i)$, a proper factorization in $\mathbf{C}[x]$.
- The polynomial $2x + 2$ is reducible over $\mathbf{Z}$ since $2x + 2 = 2(x + 1)$, a proper factorization in $\mathbf{Z}[x]$.
- We claim that the polynomial $f(x) = x^5 + 5x^2 - x - 2$ is reducible over $\mathbf{R}$. Since $f(0) = -2$ and $f(1) = 3$, and the polynomial function induced by $f(x)$ (see Exercise 9–2) is continuous, the Intermediate Value Theorem implies that $f(x)$ has a zero r (between 0 and 1). By the theorem of Section 9.4, $x - r$ divides $f(x)$, so that $f(x) = (x - r)g(x)$ for some $g(x) \in \mathbf{R}[x]$. By the theorem of Section 9.2, $g(x)$ has degree 4, so this is a proper factorization of $f(x)$.

Showing that a given polynomial is irreducible is often difficult. For irreducibility over a *field* one can sometimes use the following observation about a proper factorization, together with a proof by contradiction.

9.5.1 Theorem. *Assume that R is a field. If $f(x)$ is nonzero and it has a proper factorization $f(x) = g(x)h(x)$, then $\deg g(x), \deg h(x) < \deg f(x)$.*

Proof. Suppose that $f(x)$ is nonzero and that it has the indicated proper factorization. By Section 9.2, we have $\deg f(x) = \deg g(x) + \deg h(x)$. Now $h(x)$ is nonconstant, since it is neither zero nor a unit, so $\deg h(x) > 0$, implying

$$\deg g(x) < \deg g(x) + \deg h(x) = \deg f(x).$$

Similarly, $\deg h(x) < \deg f(x)$. □

As mentioned above, the polynomial $2x + 2$ has the proper factorization $2x+2 = 2(x+1)$ over $\mathbf{Z}$, so the theorem can fail to hold if R is not assumed to be a field.

Here are some examples to illustrate uses of the theorem:

- We claim that the polynomial $f(x) = x^2 - 2$ is irreducible over $\mathbf{Q}$. Suppose otherwise. Then it follows from the theorem that $f(x)$ has a linear factor $a_1x + a_0$. But this implies that the rational number $r = -a_0/a_1$ is a zero of $f(x)$. Now $f(x)$ has at most two zeros in $\mathbf{R}$ by Section 9.4 and we know that $\pm\sqrt{2}$ are both zeros of $f(x)$. Therefore, $r = \pm\sqrt{2}$, contradicting that $\sqrt{2}$ is irrational. We conclude that $f(x)$ is irreducible over $\mathbf{Q}$ as claimed.
- We claim that the polynomial $f(x) = x^3+x^2+2 \in \mathbf{Z}_3[x]$ is irreducible. Suppose otherwise. Since $\mathbf{Z}_3$ is a field, the theorem applies and it follows as in the preceding example that $f(x)$ has a linear factor and hence a zero $r \in \mathbf{Z}_3 = \{0, 1, 2\}$. But $f(0) = 2$, $f(1) = 1$, and $f(2) = 2$, so $f(x)$ has no zeros. We conclude that $f(x)$ is irreducible as claimed.

Other irreducibility criteria are given in Section 10.

9.6 F[x] is PID if F is field

Let F be a field.

9.6.1 Theorem. *The polynomial ring $F[x]$ is a PID.*

Proof. We first check that $F[x]$ is an integral domain. The ring $F[x]$ is commutative by Exercise 9–1. The constant polynomial 1 is an identity in $F[x]$. Let $f(x)$ and $g(x)$ be two polynomials over F. Assume that $f(x)g(x) = 0$ and $f(x) \neq 0$. By Section 9.2,

$$-\infty = \deg 0 = \deg[f(x)g(x)] = \deg f(x) + \deg g(x)$$

and, since $\deg f(x) \neq -\infty$, it must be the case that $\deg g(x) = -\infty$, that is, $g(x) = 0$. This shows that $F[x]$ has no divisors of zero. Hence $F[x]$ is an integral domain.

Let I be an ideal of $F[x]$. We claim that I is principal. If $I = 0$, then $I = (0)$ and the claim holds. Suppose that $I \neq 0$. Then I contains a nonzero polynomial, and hence a polynomial $g(x)$ of minimal degree among the nonzero elements of I.

We claim that $I = (g(x))$. The inclusion $I \supseteq (g(x))$ is immediate so we turn to the other inclusion. Let $f(x) \in I$. By the division algorithm, there exist polynomials $q(x)$ and $r(x)$ over F with $\deg r(x) < \deg g(x)$ such that $f(x) = q(x)g(x) + r(x)$. Now

$$r(x) = f(x) - q(x)g(x) \in I$$

and, since $g(x)$ has minimal degree among the nonzero elements of I, we conclude that $r(x) = 0$. Thus $f(x) = q(x)g(x) \in (g(x))$, and the equality $I = (g(x))$ is established. □

In particular, $F[x]$ is a UFD (see Section 8.5).

9.7 If R is a UFD, then so is R[x]

9.7.1 Theorem. *If R is a UFD, then the polynomial ring $R[x]$ is a UFD.*

Proof. (Omitted.) □

Here are some applications of the theorem:

- Since $\mathbf{Z}$ is a UFD, so is $\mathbf{Z}[x]$.
- A field F is a UFD, so $F[x]$ is a UFD as well. In fact, $F[x]$ is even a PID for each field F (see Section 9.6).
- If R is a UFD, then so is the **ring of polynomials over R in n indeterminants** $R[x_1, x_2, \ldots, x_n]$ ($n \in \mathbf{N}$) defined recursively by putting $R[x_1, x_2, \ldots, x_n] \cong R[x_1, x_2, \ldots, x_{n-1}][x_n]$. This claim follows immediately from the theorem by using induction on n.

9.8 Induced homomorphism of polynomial rings

Let R and R' be commutative rings with identity and let $\sigma : R \to R'$ be a homomorphism. For $f(x) \in R[x]$, denote by $\sigma f(x)$ the polynomial over R' obtained by applying σ to each coefficient of $f(x)$. In symbols, if $f(x) = \sum_{i=0}^{n} a_i x^i$, then $\sigma f(x) = \sum_{i=0}^{n} \sigma(a_i) x^i$.

9.8.1 Theorem. *The function $\bar{\sigma} : R[x] \to R'[x]$ given by $\bar{\sigma}(f(x)) = \sigma f(x)$ is a homomorphism.*

Proof. See Exercise 9–3. □

The homomorphism $\bar{\sigma} : R[x] \to R'[x]$ is the homomorphism **induced by** the homomorphism $\sigma : R \to R'$.

Let n be a positive integer. If $\sigma : \mathbf{Z} \to \mathbf{Z}_n$ is the reduction modulo n homomorphism, then the induced homomorphism $\bar{\sigma} : \mathbf{Z}[x] \to \mathbf{Z}_n[x]$ is also the **reduction modulo n homomorphism**.

For example, if $f(x) = 5x^4 + 8x^3 - 3x^2 + x - 2 \in \mathbf{Z}[x]$, then the image of $f(x)$ after applying reduction modulo 3 is $2x^4 + 2x^3 + x + 1 \in \mathbf{Z}_3[x]$.

9 – Exercises

9–1 Let R be a commutative ring with identity. Prove that $R[x]$ is commutative.

9–2 Let R be a commutative ring with identity. Define $\varphi : R[x] \to R^R$ by $\varphi(f(x))(r) = \varphi_r(f(x)) = f(r)$. Here, R^R is the ring of functions from R to R (see Section 1.5) and φ_r is the evaluation homomorphism determined by r (see Section 6.2). The function $\varphi(f(x)) : R \to R$ is the **polynomial function** induced by the polynomial $f(x)$.

(a) Prove that φ is a homomorphism.

(b) Give an example to show that φ need not be surjective.

(c) Prove that if $R = \mathbf{Z}_2$, then $R[x]/\ker\varphi \cong \mathbf{Z}_2 \oplus \mathbf{Z}_2$.

Hint: See the second paragraph of Section 1.5.

9–3 Let R and R' be commutative rings with identity and let $\sigma : R \to R'$ be a homomorphism. Prove that the map $\bar{\sigma} : R[x] \to R'[x]$ given by $\bar{\sigma}(f(x)) = \sigma f(x)$ defined in Section 9.8 is indeed a ring homomorphism as claimed.

10 Irreducibility over Q

10.1 Irreducible over Z implies irreducible over Q

Let $f(x)$ be a polynomial of degree $n > 0$ over $\mathbf{Z}$. If $f(x)$ does not factor as a product of two polynomials over $\mathbf{Z}$ each of degree strictly less than n, then it does not follow immediately that the same is true if the two polynomials are allowed to have coefficients in the larger ring $\mathbf{Q}$.

In fact, sometimes it happens that a polynomial is irreducible over one ring, but not over a larger ring. For instance, $x^2 + 1$ is irreducible over $\mathbf{R}$, but over $\mathbf{C}$ it factors as $(x + i)(x - i)$.

Nonetheless, it *is* the case that if $f(x)$ is irreducible over $\mathbf{Z}$, then it is irreducible over $\mathbf{Q}$ as well. This fact is stated in the theorem below. In order to prove the theorem we need a definition and a lemma.

The polynomial $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in \mathbf{Z}[x]$ is **primitive** if its coefficients are relatively prime, that is, if $\gcd(a_0, a_1, a_2, \ldots, a_n) = 1$.

10.1.1 Lemma (Gauss). *A product of primitive polynomials is primitive.*

Proof. We use the observation that $f(x) \in \mathbf{Z}[x]$ is primitive if and only if $\sigma_p f(x) \neq 0$ for every prime number p, where $\sigma_p : \mathbf{Z} \to \mathbf{Z}_p$ is the reduction modulo p homomorphism.

Let $f(x), g(x) \in \mathbf{Z}[x]$ be primitive polynomials and let p be an arbitrary prime number. Since $\mathbf{Z}_p[x]$ is an integral domain and $\sigma_p f(x), \sigma_p g(x) \neq 0$, we have (writing $(fg)(x)$ for $f(x)g(x)$)

$$\sigma_p(fg)(x) = \bar{\sigma}_p[f(x)g(x)] = \bar{\sigma}_p[f(x)]\bar{\sigma}_p[g(x)] = [\sigma_p f(x)][\sigma_p g(x)] \neq 0,$$

where $\bar{\sigma}_p : \mathbf{Z}[x] \to \mathbf{Z}_p[x]$ is the homomorphism induced by σ_p (see Section 9.8). Therefore, $f(x)g(x)$ is primitive. □

10.1.2 Theorem. *Let $f(x)$ be a nonconstant polynomial over $\mathbf{Z}$.*

(i) *If $f(x)$ factors over $\mathbf{Q}$ as $f(x) = g(x)h(x)$, then it factors over $\mathbf{Z}$ as $f(x) = g_1(x)h_1(x)$ with $\deg g_1(x) = \deg g(x)$ and $\deg h_1(x) = \deg h(x)$.*

(ii) *If $f(x)$ is irreducible over $\mathbf{Z}$, then it is irreducible over $\mathbf{Q}$.*

(iii) *If $f(x)$ is primitive and irreducible over* $\mathbf{Q}$*, then it is irreducible over* $\mathbf{Z}$*.*

Proof. (i) Let $f(x) = g(x)h(x)$ be a factorization of $f(x)$ with $g(x), h(x) \in \mathbf{Q}[x]$.

We may (and do) assume that the coefficients of $g(x)$ all have the same denominator $b \in \mathbf{Z}$. Let a be the greatest common divisor of the numerators of the coefficients of $g(x)$. Then $g(x) = (a/b)g'(x)$ with $g'(x)$ a primitive polynomial over $\mathbf{Z}$. Doing the same thing for the polynomial $h(x)$, multiplying the leading fractions and reducing, we conclude that

$$f(x) = (a/b)g'(x)h'(x)$$

for some primitive polynomials $g'(x), h'(x) \in \mathbf{Z}[x]$ with $\deg g'(x) = \deg g(x)$ and $\deg h'(x) = \deg h(x)$, and some relatively prime integers a and b.

We claim that $b = \pm 1$. Suppose otherwise. Then b is divisible by a prime number p. Now $bf(x) = ag'(x)f'(x)$ and, since p divides the coefficient of every term on the left, it also divides the coefficient of every term on the right. Since a and b are relatively prime, it follows that p divides the coefficient of every term of $g'(x)f'(x)$. But this product is primitive by Gauss's lemma, so this is a contradiction. Therefore, $b = \pm 1$ as claimed.

Therefore, $f(x) = g_1(x)h_1(x)$, where $g_1(x) = (a/b)g'(x) = \pm ag'(x) \in \mathbf{Z}[x]$ and $h_1(x) = h'(x) \in \mathbf{Z}[x]$. Moreover, $g_1(x)$ and $h_1(x)$ have the same degrees as $g(x)$ and $h(x)$, respectively.

(ii) We prove the contrapositive. Assume that $f(x)$ is not irreducible over $\mathbf{Q}$. Since $f(x)$ is nonconstant, it is a nonzero nonunit, so it has a proper factorization $f(x) = g(x)h(x)$ with $g(x)$ and $h(x)$ nonconstant polynomials over $\mathbf{Q}$. By part (i), $f(x) = g_1(x)h_1(x)$ with $g_1(x), h_1(x) \in \mathbf{Z}[x]$, $\deg g_1(x) = \deg g(x) > 0$, and $\deg h_1(x) = \deg h(x) > 0$. So $f(x)$ is not irreducible over $\mathbf{Z}$.

(iii) Assume that $f(x)$ is primitive and irreducible over $\mathbf{Q}$. First, $f(x)$ is a nonzero nonunit in $\mathbf{Z}[x]$ (since it is so in $\mathbf{Q}[x]$). Let $f(x) = g(x)h(x)$ be a factorization of $f(x)$ with $g(x)$ and $h(x)$ polynomials over $\mathbf{Z}$. Since $g(x)$ and $h(x)$ are polynomials over $\mathbf{Q}$ as well, and since $f(x)$ is irreducible over $\mathbf{Q}$, one or the other of these polynomials must be a unit in $\mathbf{Q}[x]$ and hence constant. We may (and do) assume that $g(x) = b_0 \in \mathbf{Z}$. Now $f(x) = b_0 h(x)$ so the coefficient of every term of $f(x)$ is divisible by b_0. Since $f(x)$ is primitive, we conclude that $g(x) = b_0 = \pm 1$, a unit in $\mathbf{Z}[x]$. This shows that $f(x)$ is irreducible over $\mathbf{Z}$. □

10.2 Rational root theorem

Let $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be a polynomial of degree $n > 0$ over $\mathbf{Z}$.

10.2.1 Theorem (Rational root theorem). *If $r \in \mathbf{Q}$ is a zero of the polynomial $f(x)$, then $r = c/d$ for some $c, d \in \mathbf{Z}$ with $c \,|\, a_0$ and $d \,|\, a_n$.*

Proof. Let $r \in \mathbf{Q}$ be a zero of $f(x)$. We can write $r = c/d$ with c and d relatively prime integers. Clearing denominators in the equation $f(c/d) = 0$ gives

$$0 = a_0 d^n + a_1 c d^{n-1} + \cdots + a_{n-1} c^{n-1} d + a_n c^n$$

so that

$$-a_0 d^n = c(a_1 d^{n-1} + \cdots + a_{n-1} c^{n-2} d + a_n c^{n-1}).$$

This shows that c divides $-a_0 d^n$. Since c and d are relatively prime, it follows that $c \,|\, a_0$. Solving the above equation for $-a_n c^n$ instead similarly reveals that $d \,|\, a_n$. □

- We claim that the polynomial $f(x) = 2 + x - 4x^2 + 3x^3$ is irreducible over $\mathbf{Q}$. If it is not irreducible, then it has a linear factor and hence a zero in $\mathbf{Q}$. Therefore, it is enough to show that it has no rational zeros. The divisors of 2 are ± 1 and ± 2 and the divisors of 3 are ± 1 and ± 3. According to the rational root theorem, the only candidates for rational zeros of $f(x)$ are

  $$\pm 1, \pm 2, \pm \frac{1}{3}, \pm \frac{2}{3}$$

  and a straightforward check shows that none of these is a zero.

10.3 Reduction modulo a prime

Let $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be a polynomial over $\mathbf{Z}$, let p be a prime number that does not divide a_n, and let $\sigma : \mathbf{Z} \to \mathbf{Z}_p$ be the reduction modulo p homomorphism. Recall that $\sigma f(x)$ denotes the polynomial over $\mathbf{Z}_p$ obtained by applying σ to each coefficient of $f(x)$.

10.3.1 Theorem. *If $\sigma f(x)$ is irreducible in $\mathbf{Z}_p[x]$, then $f(x)$ is irreducible over $\mathbf{Q}$.*

Proof. Assume that $\sigma f(x)$ is irreducible in $\mathbf{Z}_p[x]$. By way of contradiction, assume that $f(x)$ is not irreducible over $\mathbf{Q}$. Since $\sigma f(x)$ is a nonzero nonunit, it is nonconstant, so $f(x)$ is nonconstant as well and is therefore a nonzero nonunit in $\mathbf{Q}[x]$. It follows that $f(x)$ has a proper factorization and hence a factorization $f(x) = g(x)h(x)$ over $\mathbf{Q}$ with $\deg g(x), \deg h(x) < \deg f(x)$ (see Section 9.5). By (i) of Section 10.1, we may (and do) assume that $g(x), h(x) \in \mathbf{Z}[x]$.

We have $\deg \sigma f(x) = \deg f(x)$ (since $p \nmid a_n$), $\deg \sigma g(x) \leq \deg g(x)$, and $\deg \sigma h(x) \leq \deg h(x)$. Therefore, since $\sigma f(x) = [\sigma g(x)][\sigma h(x)]$,

$$\begin{aligned}\deg f(x) = \deg \sigma f(x) &= \deg \sigma g(x) + \deg \sigma h(x)\\ &\leq \deg g(x) + \deg h(x) = \deg f(x),\end{aligned}$$

and the equality of the ends forces the equalities $\deg \sigma g(x) = \deg g(x)$ and $\deg \sigma h(x) = \deg h(x)$. In particular, $\sigma g(x)$ and $\sigma h(x)$ both have degrees strictly less than the degree of $f(x)$, which is the same as the degree of $\sigma f(x)$. This shows that the factorization $\sigma f(x) = [\sigma g(x)][\sigma h(x)]$ is proper, contradicting the irreducibility of $\sigma f(x)$.

We conclude that $f(x)$ is irreducible over $\mathbf{Q}$. □

- We claim that the polynomial $f(x) = x^5 + 8x^4 + 3x^2 + 4x + 7$ is irreducible over $\mathbf{Q}$. Taking $p = 2$ in the theorem, we see that it is enough to show that $\sigma f(x) = x^5 + x^2 + 1$ is irreducible in $\mathbf{Z}_2[x]$. First, neither 0 nor 1 is a zero of $\sigma f(x)$, so this polynomial has no linear factor. Next, the only quadratic polynomials over $\mathbf{Z}_2$ are x^2, $x^2 + x$, x^2+1, and x^2+x+1. The first three of these have zeros in $\mathbf{Z}_2$ (namely, 0, 0, and 1, respectively), so none can be a factor of $\sigma f(x)$ since, as we have already seen, $\sigma f(x)$ has no zeros. Using long division, one sees that the remaining quadratic x^2+x+1 does not divide $\sigma f(x)$. Since a proper factorization of $\sigma f(x)$ would necessarily involve either a linear factor or a quadratic factor, we conclude that $\sigma f(x)$ is irreducible and the claim is established.

10.4 Eisenstein's criterion

Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ be a nonconstant polynomial over $\mathbf{Z}$ and let p be a prime number.

10.4.1 Theorem (Eisenstein's criterion). *If the following are satisfied, then $f(x)$ is irreducible over $\mathbf{Q}$:*

(i) $p^2 \nmid a_0$,

(ii) $p \,|\, a_i$ *for* $0 \leq i < n$,

(iii) $p \nmid a_n$.

Proof. Assume that the three conditions are satisfied. Suppose that $f(x)$ is not irreducible over $\mathbf{Q}$. By assumption, $f(x)$ is nonconstant and hence a nonzero nonunit in $\mathbf{Q}[x]$. Therefore, it must have a proper factorization and, using (i) of 10.1, we conclude that $f(x) = g(x)h(x)$ with $g(x)$ and $h(x)$ nonconstant polynomials over $\mathbf{Z}$. Let $\sigma : \mathbf{Z} \to \mathbf{Z}_p$ be the reduction modulo p homomorphism. By (ii), $\sigma f(x) = \sigma(a_n)x^n$ and by (iii), $\sigma(a_n)$ is nonzero and hence a unit in the field $\mathbf{Z}_p$. Now

$$\sigma(a_n)x^n = \sigma f(x) = [\sigma g(x)][\sigma h(x)],$$

and since x is an irreducible element of the UFD $\mathbf{Z}_p[x]$ (see 9.6 and 8.5), we conclude that $\sigma g(x) = r_lx^l$ and $\sigma h(x) = s_mx^m$ for some $l, m \in \mathbf{N} \cup \{0\}$ with r_l and s_m nonzero elements of $\mathbf{Z}_p$. Since $l \leq \deg g(x)$ and $m \leq \deg h(x)$, we have, using equations from above,

$$n = l + m \leq \deg g(x) + \deg h(x) = \deg f(x) = n,$$

which forces $l = \deg g(x) > 0$ and $m = \deg h(x) > 0$. In particular, the constant terms of $\sigma g(x)$ and $\sigma h(x)$ are both zero. This, in turn, implies that the constant terms of $g(x)$ and $h(x)$ are both divisible by p. However, the constant term a_0 of $f(x)$ is the product of these latter constant terms, so we get $p^2 \,|\, a_0$ in opposition to (i).

We conclude that $f(x)$ is irreducible over $\mathbf{Q}$ as claimed. □

- The polynomial $2x^5 - 6x^3 + 9x^2 - 15 \in \mathbf{Z}[x]$ is irreducible over $\mathbf{Q}$ by Eisenstein's criterion with $p = 3$.

10.5 Shift of indeterminate

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial over $\mathbf{Z}$ and let c be an integer. Denote by $f(x+c)$ the polynomial obtained from $f(x)$ by replacing each occurrence of the indeterminate x by the "shifted indeterminate" $x+c$, that is, $f(x+c) = \sum_{i=0}^{n} a_i (x+c)^i$.

10.5.1 Theorem. *If $f(x+c)$ is irreducible over $\mathbf{Z}$, then $f(x)$ is irreducible over $\mathbf{Z}$ and is therefore irreducible over $\mathbf{Q}$ if it is nonconstant.*

Proof. Assume that $f(x+c)$ is irreducible over $\mathbf{Z}$. The map $\varphi_{x+c} : \mathbf{Z}[x] \to \mathbf{Z}[x]$ given by $\varphi_{x+c}(f(x)) = f(x+c)$ is an isomorphism (the proof that the evaluation map is a homomorphism easily generalizes to show that this map is a homomorphism as well, and this map is bijective since φ_{x-c} is an inverse). Since $f(x+c)$ is irreducible over $\mathbf{Z}$ (which means after all that it is an irreducible element of the ring $\mathbf{Z}[x]$), and since it corresponds to $f(x)$ under the isomorphism, we conclude that $f(x)$ is also irreducible over $\mathbf{Z}$.

Finally, assuming $f(x)$ is nonconstant, it follows from (ii) of 10.1 that $f(x)$ is irreducible over $\mathbf{Q}$. □

- Let p be a prime number. The polynomial $f(x) = (x^p - 1)/(x-1) = x^{p-1} + x^{p-2} + \cdots + x + 1 \in \mathbf{Z}[x]$ is the p **th cyclotomic polynomial**. We claim that $f(x)$ is irreducible over $\mathbf{Q}$. By the theorem, it suffices to show that the polynomial $f(x+1)$ is irreducible over $\mathbf{Z}$. Using the binomial theorem, we get

$$\begin{aligned} f(x+1) &= \frac{(x+1)^p - 1}{(x+1) - 1} \\ &= \frac{(x^p + \binom{p}{1}x^{p-1} + \cdots + \binom{p}{p-1}x + 1) - 1}{x} \\ &= x^{p-1} + \binom{p}{1}x^{p-2} + \cdots + \binom{p}{p-2}x + \binom{p}{p-1}. \end{aligned}$$

  Now p divides each binomial coefficient $\binom{p}{i}$ with $0 < i < p$ and $p^2 \nmid p = \binom{p}{p-1}$, so $f(x+1)$ is irreducible over $\mathbf{Q}$ by Eisenstein's criterion. Since the coefficient of x^{p-1} is one, the polynomial $f(x+1)$ is primitive, so it is irreducible over $\mathbf{Z}$ as well by (iii) of 10.1. This establishes the claim.

10 – Exercises

10–1 Prove that the polynomial $f(x) = 3x^4 + 8x^3 + 8x^2 - 2x - 3$ is irreducible over $\mathbf{Q}$.

Hint: Consider $f(x-1)$.

10–2 Prove that the polynomial $f(x) = 3x^2 - 7x - 5$ is irreducible over $\mathbf{Q}$.

11 Vector space

11.1 Definition

Let F be a field. A **vector space** over F is a triple $(V, +, \cdot\,)$, where $(V, +)$ is an abelian group and $\cdot$ is a function $F \times V \to V$ (written $(a, v) \mapsto av$) satisfying the following for all $a, b \in F$ and all $v, w \in V$:

(i) $a(v + w) = av + aw$,

(ii) $(a + b)v = av + bv$,

(iii) $a(bv) = (ab)v$,

(iv) $1v = v$.

Let $(V, +, \cdot\,)$ be a vector space over F. The elements of V are called **vectors**, the elements of F are called **scalars**, and the function $\cdot$ is called **scalar multiplication**. Parts (i) and (ii) are both **distributive properties**. Part (iii) is the **associative property** of scalar multiplication. If the operations are clear from context, we say that V is a vector space over F.

11.2 Examples

- Let F be a field and let n be a positive integer. The set F^n of n-tuples of elements of F is a vector space over F with componentwise addition and with scalar multiplication defined by
$$a(a_1, a_2, \ldots, a_n) = (aa_1, aa_2, \ldots, aa_n).$$

In particular, $\mathbf{R}^n$ is a vector space over $\mathbf{R}$.

- Let F be a field. The polynomial ring $F[x]$ is a vector space over F with addition being addition of polynomials and with scalar multiplication being given by

  $$a(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = aa_0 + aa_1x + aa_2x^2 + \cdots + aa_nx^n.$$

- Let F be a field and let n be a positive integer. The set $\mathrm{Mat}_n(F)$ of $n \times n$ matrices over F is a vector space with addition being matrix addition and scalar multiplication being defined by

  $$a[a_{ij}] = [aa_{ij}]$$

- Let E be a field and let F be a subfield of E (meaning that F is a subring of E and it is also a field). It follows immediately from the ring axioms that E is a vector space over F if we take as addition the addition in E and as scalar multiplication the product in E.

- Let F be a field. An **algebra** over F is a pair (R, φ) where R is a ring with identity and $\varphi : F \to R$ is a homomorphism such that $\varphi(1) = 1$ and $\operatorname{im} f$ is contained in the center of R.

  Let (R, φ) be an F-algebra. Then R is a vector space over F with addition being that in R and scalar multiplication being defined by $ar = \varphi(a)r$. The examples above are all algebras with φ given, respectively, as follows:

  $$\begin{array}{lll} \mathbf{R} \to \mathbf{R}^n & \text{by} & a \mapsto (a, \ldots, a), \\ F \to F[x] & \text{by} & a \mapsto a \text{ (constant polynomial)}, \\ F \to \mathrm{Mat}_n(F) & \text{by} & a \mapsto aI \text{ (scalar matrix)}, \\ F \to E & \text{by} & a \mapsto a. \end{array}$$

  Since φ is nonzero, its kernel is trivial (being a proper ideal in the field F), so that φ is injective. Therefore, F is isomorphic to its image under φ, which is a subring of R containing the element 1. It is convenient to use this isomorphism to view F as a subring of R.

11.3 Basic identities

Let F be a field and let V be a vector space over F.

11.3.1 Theorem.

(i) $0v = 0$ *for all* $v \in V$,

(ii) $a0 = 0$ *for all* $a \in F$,

(iii) $(-a)v = -av$ *and* $a(-v) = -av$ *for all* $a \in F, v \in V$.

Proof. (i) Let $v \in V$. Since $0v + 0v = (0 + 0)v = 0v$, cancellation gives $0v = 0$.

(ii) Let $a \in F$. Since $a0 + a0 = a(0 + 0) = a0$, cancellation gives $a0 = 0$.

(iii) Let $a \in F$ and $v \in V$. Since $av + (-a)v = (a + (-a))v = 0v = 0$ (using part (i)), we have $(-a)v = -av$.

Since $av + a(-v) = a(v + (-v)) = a0 = 0$ (using part (ii)), we have $a(-v) = -av$ □

11.4 Subspace

Let F be a field and let V be a vector space over F. A **subspace** of V is a subgroup of $(V, +)$ that is closed under scalar multiplication. Thus, a subset W of V is a subspace if and only if

(i) $0 \in W$,

(ii) $w, w' \in W \Rightarrow w + w' \in W$,

(iii) $w \in W \Rightarrow -w \in W$,

(iv) $a \in F, w \in W \Rightarrow aw \in W$.

We write $W \leq V$ to indicate that W is a subspace of V. Both V and $\{0\}$ are subspaces of V.

- Let n be a positive integer. For each $1 \leq k \leq n$, the set

  $$\{(0, \ldots 0, \underset{k}{a}, 0, \ldots, 0) \,|\, a \in F\}$$

  is a subspace of F^n.

- Let n be a nonnegative integer. The set of all polynomials over F of degree $\leq n$ is a subspace of $F[x]$.
- Let n be a positive integer. The set
$$\{[a_{ij}] \,|\, a_{ij} = 0 \text{ for all } i > j\}$$
of all upper triangular $n \times n$ matrices over F is a subspace of $\mathrm{Mat}_n(F)$.

11.5 Quotient space

Let V be a vector space and let W be a subspace of V. Then W is a subgroup of the additive group $(V, +)$ and, since this latter group is abelian, W is normal. Therefore, the quotient group V/W is defined and it is abelian. In fact, the naturally defined scalar multiplication makes this group into a vector space.

In more detail, $V/W = \{v + W \,|\, v \in V\}$ is a vector space with addition and scalar multiplication given by

(i) $(v + W) + (v' + W) = (v + v') + W$,

(ii) $a(v + W) = (av) + W$.

These operations are well-defined, meaning that they do not depend on the choices of coset representatives. V/W is the **quotient** of V by W.

11.6 Span

Let F be a field, let V be a vector space over F, and let S be a subset of V. The **span** of S, written $\langle S \rangle$, is the intersection of all subspaces of V that contain S:
$$\langle S \rangle = \bigcap_{\substack{W \leq V \\ W \supseteq S}} W.$$
Since an intersection of subspaces is again a subspace, the span of S is a subspace of V. It is the smallest subspace of V containing S in the sense that if W is a subspace of V containing S, then $\langle S \rangle \subseteq W$.

The subset S **spans** V if $\langle S \rangle = V$. To show S spans V it is enough to show $V \subseteq \langle S \rangle$ since the other inclusion always holds.

Assume $S = \{v_1, v_2, \ldots, v_n\}$, a finite set. We write $\langle v_1, v_2, \ldots, v_n\rangle$ instead of $\langle \{v_1, v_2, \ldots, v_n\}\rangle$ and call it the span of the vectors $v_1, v_2, \ldots, v_n$.

11.6.1 Theorem. $\langle v_1, v_2, \ldots, v_n\rangle = \{a_1v_1 + a_2v_2 + \cdots + a_nv_n \,|\, a_i \in F\}$.

Proof. One checks that the right-hand side is a subspace of V. It contains each v_i (since one can let $a_i = 1$ and $a_j = 0$, $j \neq i$), so it contains the left-hand side. On the other hand, every subspace of V containing each v_i must contain the right-hand side by the closure properties, so the left-hand side contains the right-hand side. Therefore, the stated equality holds. □

The elements in the set on the right are **linear combinations** of the vectors $v_1, v_2, \ldots, v_n$.

- Let $v_1 = (1, 0, 0)$ and $v_2 = (0, 1, 0)$ in $\mathbf{R}^3$. Then

  $$\langle v_1, v_2\rangle = \{(x, y, 0) \,|\, x, y \in \mathbf{R}\},$$

  the xy-plane.

- Let $v_1 = (1, 1)$ and $v_2 = (-1, 1)$ in $\mathbf{R}^2$. Then the set $S = \{v_1, v_2\}$ spans $\mathbf{R}^2$. Indeed, if $(x, y) \in \mathbf{R}^2$, then $(x, y) = a_1v_1 + a_2v_2 \in \langle S\rangle$, where $a_1 = (x + y)/2$ and $a_2 = (y - x)/2$, whence $\mathbf{R}^2 \subseteq \langle S\rangle$.

11.7 Linear independence

Let F be a field, let V be a vector space over F, and let $S = \{v_1, v_2, \ldots, v_n\}$ be a finite subset of V. The set S is **linearly independent** if

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0 \quad \Rightarrow \quad a_i = 0 \text{ for all } i$$

($a_i \in F$). In other words, the set S is linearly independent if and only if the only way to get a linear combination of the vectors in S to equal the zero vector is to use all zero coefficients.

If S is not linearly independent, then it is **linearly dependent**. Thus, S is linearly dependent if and only if there exists a linear combination of the vectors in S equaling zero with coefficients *not all* zero.

Sometimes, instead of saying the set $S = \{v_1, v_2, \ldots, v_n\}$ is linearly independent, we say that the vectors $v_1, v_2, \ldots, v_n$ are linearly independent (and similarly for linear dependence).

- We claim that the vectors $v_1 = (1,1)$ and $v_2 = (-1,1)$ of $\mathbf{R}^2$ are linearly independent. Suppose we have $a_1v_1 + a_2v_2 = 0$ ($a_i \in \mathbf{R}$). Then
$$(0,0) = a_1(1,1) + a_2(-1,1) = (a_1 - a_2, a_1 + a_2)$$
implying $a_1 - a_2 = 0$ and $a_1 + a_2 = 0$. Solving this system yields $a_1 = 0$ and $a_2 = 0$. This establishes the claim.

- We claim that S is linearly dependent if it contains the zero vector. We assume, without loss of generality, that $v_1 = 0$. Then
$$1v_1 + 0v_2 + 0v_3 + \cdots + 0v_n = 0.$$
Since not all coefficients are zero, the claim is established.

11.8 Basis

Let F be a field and let V be a vector space over F. A **basis** for V is a subset of V that spans V and is linearly independent. Thus, the subset $S = \{v_1, v_2, \ldots, v_n\}$ of V is a basis for V if and only if

(i) $V = \langle v_1, v_2, \ldots, v_n \rangle$,

(ii) $v_1, v_2, \ldots, v_n$ are linearly independent.

According to Section 11.6, property (i) is equivalent to the statement that every $v \in V$ can be written in the form $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ for some $a_i \in F$.

The plural of basis is “bases”.

- Let $e_1 = (1,0)$ and $e_2 = (0,1)$. The set $\{e_1, e_2\}$ spans $\mathbf{R}^2$ and it is linearly independent, so it is a basis for $\mathbf{R}^2$. More generally, if n is a positive integer, then the set $\{e_1, e_2, \ldots, e_n\}$, where $e_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0)$, is a basis for F^n, the **standard basis**.

- In view of the examples in the last two sections, the set $\{v_1, v_2\}$, where $v_1 = (1,1)$ and $v_2 = (-1,1)$, is a basis for $\mathbf{R}^2$.

These examples show that there can be more than one basis for a vector space. It will be shown, however, that any two bases for a given vector space must have the same number of elements (Section 11.9).

Let $S = \{v_1, v_2, \dots, v_n\}$ be a finite subset of V.

11.8.1 Theorem. *The set S is a basis for V if and only if every vector v in V can be written uniquely in the form $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ with $a_i \in F$.*

Proof. Assume that S is a basis for V. Let $v \in V$. By the spanning property of basis, v can be written in at least one way in the form $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ with $a_i \in F$. Suppose also $v = b_1v_1 + b_2v_2 + \cdots + b_nv_n$ with $b_i \in F$. Then the two linear combinations are equal, so rearranging we get

$$(a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_n - b_n)v_n = 0.$$

By the linear independence property of basis we get $a_i - b_i = 0$ for all i, so that $a_i = b_i$ for all i. This demonstrates the uniqueness claim.

Now assume that every vector v in V can be written uniquely in the form $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ with $a_i \in F$. It is immediate that S spans V. Suppose we have $a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0$ with $a_i \in F$. We also have $0v_1 + 0v_2 + \cdots + 0v_n = 0$, so these two ways of writing the zero vector 0 must be the same by the uniqueness assumption. Therefore, $a_i = 0$ for all i, showing that S is linearly independent. We conclude that S is a basis. □

11.9 Dimension

Let F be a field and let V be a vector space over F.

11.9.1 Theorem.

(i) *Every finite spanning subset of V contains a basis for V.*

(ii) *Assume V has a spanning set consisting of n vectors. Every linearly independent subset of V is contained in a basis for V consisting of at most n vectors.*

(iii) *Any two bases for V have the same number of elements.*

Proof. (i) (Sketch) If S is a finite spanning subset of V, then a subset of S chosen to be minimal among all subsets of S that span V is a basis.

(ii) Let $v_1, v_2, \dots, v_m$ be linearly independent vectors in V. We prove that $m \leq n$ arguing by induction on n. If $n = 0$, then $V = \{0\}$ forcing $m = 0 \leq n$ since no set of vectors in $\{0\}$ is linearly independent.

Assume that $n > 0$. By assumption, V has a spanning set $S = \{w_1, w_2, \dots, w_n\}$ consisting of n vectors. For each $1 \leq i \leq m$ there exist scalars a_{ij}, $1 \leq j \leq n$, such that $v_i = \sum_j a_{ij} w_j$. If $a_{i1} = 0$ for each i, then the v_i are contained in $\langle w_2, w_3, \dots, w_n \rangle$ so that $m \leq n - 1 < n$ by the induction hypothesis. Therefore, by renumbering if necessary, we may (and do) assume that $a_{11} \neq 0$. Define $v_i' = v_i - a_{i1} a_{11}^{-1} v_1 \in \langle w_2, \dots, w_n \rangle$. Then $v_2', \dots, v_m'$ are linearly independent. By the induction hypothesis, $m - 1 \leq n - 1$, so $m \leq n$ as desired.

Now we prove the claim in the statement of the theorem. By (i) we may (and do) assume that V has a basis B having n elements. Let S' be a linearly independent subset of V. If $S \subseteq B$ is chosen to be maximal among all subsets of B such that $S' \cup S$ is linearly independent, then this union is a basis, and it consists of at most n vectors by the previous paragraph.

(iii) Use (ii). □

The vector space V is **finite dimensional** of **dimension** n, written $\dim V = n$, if it has a basis consisting of n vectors (well-defined by (iii)). If V does not have a (finite) basis, it is **infinite dimensional**.

- For a positive integer n, the vector space $\mathbf{R}^n$ has dimension n since the standard basis has n elements.
- For a positive integer n, the vector space consisting of all polynomials over F of degree less than n has dimension n, since $\{1, x, x^2, \dots, x^{n-1}\}$ is a basis, as is easily checked.

We gather together some useful consequences of the theorem:

11.9.2 Corollary.

(i) *V is finite dimensional if and only if it has a finite spanning set.*

(ii) *If V is finite dimensional, then a subset of V is a basis for V if and only if it is a minimal spanning subset of V*

(iii) *If V is finite dimensional, then a subset of V is a basis for V if and only if it is a maximal linearly independent subset of V.*

(iv) *If V is finite dimensional of dimension n, then a subset of V consisting of n vectors is a basis for V if either it spans V or it is linearly independent.*

(v) *If V is finite dimensional and W is a subspace of V, then W is finite dimensional,* $\dim W \leq \dim V$*, and* $\dim V/W = \dim V - \dim W$.

11.10 Linear transformation

Let F be a field and let V and V' be vector spaces over F. A **linear transformation** from V to V' is a function $T : V \to V'$ satisfying the following:

(i) $T(v + w) = T(v) + T(w)$ for all $v, w \in V$,

(ii) $T(av) = aT(v)$ for all $a \in F$, $v \in V$.

An **isomorphism** from V to V' is a bijective linear transformation from V to V'. The vector spaces V and $V'$are **isomorphic**, written $V \cong V'$, if there exists an isomorphism from V to V'. Two isomorphic vector spaces are identical as far as their vector properties are concerned.

11.10.1 Theorem. *A vector space over F of finite dimension n is isomorphic to F^n.*

Proof. (Sketch) Let V be a vector space over F of dimension n. There exists a basis of V having n elements, say, $S = \{v_1, v_2, \ldots, v_n\}$. The map $T : F^n \to V$ given by

$$T((a_1, a_2, \ldots, a_n)) = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

is a linear transformation. It is bijective by the theorem of Section 11.8. □

11.11 First Isomorphism Theorem

Let $T : V \to V'$ be a linear transformation. Part (i) of 11.10 says that T is a group homomorphism from $(V, +)$ to $(V', +)$. From group theory, we know that the kernel of T is a subgroup of V and the image of T is a subgroup of V'. These subgroups are closed under scalar multiplication so that they are in fact both subspaces.

The kernel of T is sometimes called the **null space** of T and the image of T is sometimes called the **range** of T.

11.11.1 Theorem (First Isomorphism Theorem).

$$V/\ker T \cong \operatorname{im} T.$$

Proof. Define $\bar{T} : V/\ker T \to \operatorname{im} T$ by $\bar{T}(v + \ker T) = T(v)$. Since T is a group homomorphism, this is a well-defined isomorphism of groups by the proof of the first isomorphism theorem for groups. Since

$$\bar{T}(a(v + \ker T)) = \bar{T}((av) + \ker T) = T(av) = aT(v) = a\bar{T}(v + \ker T)$$

for every $a \in F$ and every $v \in V$, $\bar{T}$ is a linear transformation and hence an isomorphism of vector spaces as well. □

Assume that V is finite dimensional. The image of T is a finite-dimensional since it is spanned by the image of a basis for V (use Section 11.9 Corollary (i)). Its dimension is the **rank** of T, written $\operatorname{rank} T$. The kernel of T is also finite-dimensional (Section 11.9 Corollary (v)). Its dimension is the **nullity** of T, written $\operatorname{nullity} T$. In symbols,

- $\operatorname{rank} T = \dim \operatorname{im} T$,
- $\operatorname{nullity} T = \dim \ker T$.

The following result is an immediate consequence of the First Isomorphism Theorem.

11.11.2 Corollary (Rank plus nullity theorem).

$$\dim V = \operatorname{rank} T + \operatorname{nullity} T$$

- Let $T : \mathbf{R}^3 \to \mathbf{R}^3$ be "projection onto the (x, y)-plane." That is, T is the function given by $T(x, y, z) = (x, y, 0)$. Then T is a linear transformation. Its image is the (x, y)-plane, which has dimension two, and its kernel is the z-axis, which has dimension one. We have

  $$\dim \mathbf{R}^3 = 3 = 2 + 1 = \dim \operatorname{im} T + \dim \ker T = \operatorname{rank} T + \operatorname{nullity} T$$

  in agreement with the corollary.

  In more detail, the quotient space $\mathbf{R}^3/\ker T$ is the set of all cosets of the z-axis, that is, it is the set of all lines in $\mathbf{R}^3$ that are parallel to the z-axis. There is a natural bijection from this set of lines to the

(x, y)-plane; it is defined by sending a line to its intersection with the (x, y)-plane. This bijection is compatible with the operations (i.e., it is a linear transformation), and it is therefore an isomorphism. This illustrates the First Isomorphism Theorem.

11 – Exercises

11–1 Let V be a vector space over a field F and let $S = \{v_1, v_2, \ldots, v_n\}$ be a linearly independent subset of V. Prove that every subset of S is linearly independent.

11–2 Let V be the subspace of $\mathbf{R}[x]$ consisting of all polynomials over $\mathbf{R}$ of degree at most two. Prove that the set $S = \{1 + x, x - 2x^2, 1 + 3x^2\}$ is a basis for V.

11–3 Let F be a field, let V be a vector space over F, and let $S = \{v_1, v_2, \ldots, v_n\}$ be a subset of V.

(a) Prove that S is linearly dependent if and only if one of the vectors v_i is in the span of the other vectors.

(b) Prove or give a counterexample: If S is linearly dependent, then $v_1 \in \langle v_2, v_3, \ldots, v_n\rangle$.

11–4 Let V be a vector space over a field F and let $B = (v_1, v_2, \ldots, v_n)$ be an *ordered* basis for V (so $\{v_1, v_2, \ldots, v_n\}$ is a basis for V and the basis vectors have the indicated fixed ordering).

Let $v \in V$. By 11.8, $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ for uniquely determined $a_i \in F$. The **coordinate vector** of v relative to B, denoted $[v]_B$, is the $n \times 1$ matrix defined by

$$[v]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Let V' be another vector space over F with ordered basis $B' = (v_1', v_2', \ldots, v_m')$ and let $T : V \to V'$ be a linear transformation. The **matrix of** T relative to the bases B and B' is the $m \times n$ matrix A having jth column $[T(v_j)]_{B'}$ $(1 \leq j \leq n)$.

(a) Prove that $[T(v)]_{B'} = A[v]_B$ for each $v \in V$.

(b) Let V be the vector space of polynomials over $\mathbf{R}$ of degree at most two. Then $B = (1+x, x-2x^2, 1+3x^2)$ is an ordered basis for V (c.f. Exercise 11–2) and so also is $B' = (x^2, x, 1)$. The function $T : V \to V$ defined by $T(f(x)) = f'(x)$ (= derivative of $f(x)$) is a linear transformation. Find the matrix of T relative to B and B' and verify that the formula of part (a) is valid.

12 Field extension

12.1 Definition of field extension and degree

Let E be a field and let F be a subring of E that is also a field. We say that F is a **subfield** of E and that E is a **field extension** of F. For brevity, we will refer to this situation by saying that $E \supseteq F$ is a field extension.

Some examples of field extensions are $\mathbf{R} \supseteq \mathbf{Q}$ and $\mathbf{C} \supseteq \mathbf{R}$.

Essential to our discussion is the simple observation that E can be regarded as a vector space over F with both addition and scalar multiplication coming from the operations in E. The dimension of this vector space is written $[E : F]$ and is called the **degree** of the field extension $E \supseteq F$. A field extension is **finite** or **infinite** according as its degree is finite or infinite.

- Every complex number can be written uniquely in the form $a+bi$ with $a, b \in \mathbf{R}$, so $\{1, i\}$ is a basis for the vector space $\mathbf{C}$ over $\mathbf{R}$. Therefore, $[\mathbf{C} : \mathbf{R}] = 2$.

- We will see that $\mathbf{R} \supseteq \mathbf{Q}$ is an infinite extension.

12.2 Degree is multiplicative

Let $E \supseteq M \supseteq F$ be field extensions (sometimes referred to as a **tower** of fields).

12.2.1 Theorem. *The degree $[E : F]$ is finite if and only if both of the degrees $[E : M]$ and $[M : F]$ are finite, and in this case*

$$[E : F] = [E : M][M : F].$$

Proof. Assume that the degree $[E : F]$ is finite. Then E, viewed as a vector space over F, has a (finite) basis B. Since E is the span of B over F, it is the span of B over M as well (using that $M \supseteq F$), so $[E : M]$ is finite by Section 11.9. Also, M is a subspace of the finite dimensional vector space E over F, so its dimension $[M : F]$ is finite as well (part (v) of the corollary in Section 11.9).

Now assume that $[E : M]$ and $[M : F]$ are both finite. Then E has a basis $B = \{\beta_1, \ldots, \beta_n\}$ over M and M has a basis $A = \{\alpha_1, \ldots, \alpha_m\}$ over F. We claim that the set $AB = \{\alpha_i\beta_j \,|\, 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis for E over F.

Let $c \in E$. Then $c = \sum_j b_j\beta_j$ for some $b_j \in M$. In turn, for each j, $b_j = \sum_i a_{ij}\alpha_i$ for some $a_{ij} \in F$. Therefore,

$$c = \sum_j b_j\beta_j = \sum_j \left(\sum_i a_{ij}\alpha_i\right)\beta_j = \sum_{i,j} a_{ij}\alpha_i\beta_j,$$

so AB spans E over F.

Now suppose that $\sum_{i,j} a_{ij}\alpha_i\beta_j = 0$ with $a_{ij} \in F$. Then rearranging the sum, we get $\sum_j \left(\sum_i a_{ij}\alpha_i\right)\beta_j = 0$, so linear independence of B gives $\sum_i a_{ij}\alpha_i = 0$ for all j. In turn, linear independence of A gives $a_{ij} = 0$ for all i and j. Therefore, AB is linearly independent.

We conclude that AB is a basis for E over F and $[E : F] = |AB| = |B||A| = [E : M][M : F]$. □

12.3 Subfield generated by a set

Let E be a field. If S a subset of E, then the intersection of all subfields of E containing S is a subfield of E called the **subfield** of E **generated**

by S. It is the smallest subfield of E that contains S in the sense that it is contained in every subfield of E that contains S.

Let $E \supseteq F$ be a field extension and let S be a subset of E. We write $F(S)$ for the subfield of E generated by $F \cup S$. If $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, then we write $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ for $F(S)$ and refer to this as the field obtained from F by adjoining $\alpha_1, \alpha_2, \dots, \alpha_n$.

The extension $E \supseteq F$ is **simple** if $E = F(\alpha)$ for some $\alpha \in E$.

- The extension $\mathbf{C} \supseteq \mathbf{R}$ is simple since $\mathbf{C} = \mathbf{R}(i)$.

12.4 Algebraic element

Let $E \supseteq F$ be a field extension and let $\alpha \in E$. The element α is **algebraic over** F if α is a zero of a nonzero polynomial over F. So α is algebraic over F if there exists $f(x) \in F[x]$ with $f(x) \neq 0$ such that $f(\alpha) = 0$. If α is not algebraic, it is **transcendental**.

- The real number $\sqrt{2}$ is algebraic over $\mathbf{Q}$, since $f(\sqrt{2}) = 0$, where $f(x) = x^2 - 2 \in \mathbf{Q}[x]$.
- The complex number i is algebraic over $\mathbf{R}$, since $f(i) = 0$, where $f(x) = x^2 + 1 \in \mathbf{R}[x]$. Since $f(x) = x^2 + 1$ is also a polynomial over $\mathbf{Q}$, the number i is algebraic over $\mathbf{Q}$ as well.
- If α is in F, then α is algebraic over F, since $f(\alpha) = 0$, where $f(x) = x - \alpha \in F[x]$.
- It has been shown that the real numbers π and e are both transcendental over $\mathbf{Q}$.

Assume that α is also an element of another field extension E' of F. In posing the question of whether a is algebraic over F, it does not matter whether one views α as an element of E or as an element of E'. Indeed, if $f(x)$ is a polynomial over F, then $f(\alpha)$ is an element of $F(\alpha)$, which is a subfield of both E and E', so the answer to the question of whether $f(\alpha)$ is zero does not depend on the choice of E or E' as the field extension.

A polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$ over F with $a_n = 1$ is a **monic polynomial**.

12.4.1 Theorem. *Assume that α is algebraic over F.*

(i) *There exists a unique monic polynomial $p_\alpha(x)$ over F of minimal degree such that $p_\alpha(\alpha) = 0$.*

(ii) *For every $f(x) \in F[x]$, we have $f(\alpha) = 0$ if and only if $p_\alpha(x)$ divides $f(x)$.*

(iii) *$p_\alpha(x)$ is irreducible.*

(iv) *$F(\alpha) \cong F[x]/(p_\alpha(x))$ and $F(\alpha) = \{f(\alpha) \mid f(x) \in F[x]\}$.*

(v) *Put $n = \deg p_\alpha(x)$. The set $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is a basis for the vector space $F(\alpha)$ over F.*

(vi) *$[F(\alpha) : F] = \deg p_\alpha(x)$.*

Proof. (i), (ii) Since α is algebraic over F, it is a zero of a nonzero polynomial over F. From among all such polynomials, choose one with least degree and call it $p_\alpha(x)$. Since α is a zero of any scalar multiple of $p_\alpha(x)$, we may (and do) assume that $p_\alpha(x)$ is monic.

Let $f(x)$ be a polynomial over F. Assume that $f(\alpha) = 0$. By the division algorithm, there exist $q(x), r(x) \in F[x]$ with $\deg r(x) < \deg p_\alpha(x)$ such that $f(x) = q(x)p_\alpha(x) + r(x)$. Since $r(\alpha) = f(\alpha) - q(\alpha)p_\alpha(\alpha) = 0$, the choice of $p_\alpha(x)$ forces $r(x) = 0$, whence $f(x) = q(x)p_\alpha(x)$. Therefore, $p_\alpha(x)$ divides $f(x)$. On the other hand, if $p_\alpha(x)$ divides $f(x)$, then $f(x) = g(x)p_\alpha(x)$ for some $g(x) \in F[x]$, so $f(\alpha) = g(\alpha)p_\alpha(\alpha) = 0$. This proves (ii).

Let $f(x)$ be a polynomial over F with $f(\alpha) = 0$. By (ii), $f(x) = g(x)p_\alpha(x)$ for some $g(x) \in F[x]$. If $\deg f(x) = \deg p_\alpha(x)$, then $g(x)$ is a constant polynomial, and if, additionally, $f(x)$ is monic, then $g(x) = 1$, so that $f(x) = p_\alpha(x)$. This establishes the uniqueness statement in (i) and finishes the proof of that part.

(iii) Assume, to the contrary, that $p_\alpha(x)$ is not irreducible. First, $p_\alpha(x)$ is nonzero (since it is monic) and a nonunit (since having α as a zero means it is nonconstant). Therefore, there is a proper factorization $p_\alpha(x) = f(x)g(x)$. By Section 9.5, we have $\deg f(x), \deg g(x) < \deg p_\alpha(x)$. Now $f(\alpha)g(\alpha) = p_\alpha(\alpha) = 0$ and, since F is an integral domain, we get $f(\alpha) = 0$ or $g(\alpha) = 0$. Either case contradicts the minimality of the degree of $p_\alpha(x)$. Therefore, $p_\alpha(x)$ is irreducible.

(iv) The kernel I of the evaluation homomorphism $\varphi_\alpha : F[x] \to E$ (which sends $f(x)$ to $f(\alpha)$) is the set of all $f(x) \in F[x]$ for which $f(\alpha) = 0$. In view of part (ii), we have $I = (p_\alpha(x))$, the principal ideal generated by $p_\alpha(x)$. Since $p_\alpha(x)$ is irreducible (by (iii)), the ideal I is maximal (8.3 and 9.6), so the quotient ring $F[x]/I$ is a field (5.5). Therefore, the image

$$M = \{f(\alpha) \mid f(x) \in F[x]\}$$

of φ_α, which is isomorphic to $F[x]/I$ by the first isomorphism theorem, is a subfield of E.

We claim that $M = F(\alpha)$. Any subfield of E that contains F and α must contain M by the closure properties, so $M \subseteq F(\alpha)$. On the other hand, M is a subfield of E that contains each $a \in F$ (let $f(x) = a$) and also α (let $f(x) = x$), so $M \supseteq F(\alpha)$. This establishes the claim and finishes the proof of (iv).

(v) Put $S = \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$, where $n = \deg p_\alpha(x)$. Since α is an element of $F(\alpha)$, it follows from closure properties that $S \subseteq F(\alpha)$. Let $\beta \in F(\alpha)$. By part (iv), we have $\beta = f(\alpha)$ for some $f(x) \in F[x]$. By the division algorithm, there exist $q(x), r(x) \in F[x]$ with $\deg r(x) < \deg p_\alpha(x)$ such that $f(x) = q(x)p_\alpha(x) + r(x)$, so

$$\beta = f(\alpha) = q(\alpha)p_\alpha(\alpha) + r(\alpha) = r(\alpha).$$

Now $\deg r(x) < \deg p_\alpha(x) = n$, so $r(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$ for some $a_i \in F$. Therefore,

$$\beta = r(\alpha) = a_0 \cdot 1 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} \in \langle S \rangle.$$

This shows that S spans $F(\alpha)$. Suppose that $a_0 \cdot 1 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} = 0$, with $a_i \in F$. Then $g(x) = \sum_i a_ix^i$ is a polynomial over F of degree less than $\deg p_\alpha(x)$ such that $g(\alpha) = 0$. By part (i), $g(x) = 0$, that is, $a_i = 0$ for all i. Thus S is linearly independent. We conclude that S is a basis for $F(\alpha)$ as claimed.

(vi) This follows immediately from part (v) □

The polynomial $p_\alpha(x)$ is the **minimal polynomial** of α over F. Generally, the base field F of the extension $E \supseteq F$ will be fixed in our discussion, so the notation $p_\alpha(x)$ will be unambiguous. However, when we need to reference the base field in the notation we will write $p_{\alpha,F}(x)$.

- As was observed earlier, the set $\{1, i\}$ is a basis for the vector space $\mathbf{C} = \mathbf{R}(i)$ over $\mathbf{R}$. This is in agreement with the theorem since $p_i(x) = x^2 + 1$ is the minimal polynomial of i over $\mathbf{R}$.

12.5 A finite extension is algebraic

Let $E \supseteq F$ be a field extension. If every element of E is algebraic over F then $E \supseteq F$ is an **algebraic extension**.

12.5.1 Theorem. *If $E \supseteq F$ is finite, then it is algebraic.*

Proof. Assume that $E \supseteq F$ is finite, so that $[E : F] = n$ for some positive integer n. Let α be an element of E. Since E has dimension n over F, the $n + 1$ elements $1, \alpha, \alpha^2, \dots, \alpha^n$ must be linearly dependent. So there exist $a_i \in F$, not all zero, such that

$$a_0 \cdot 1 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n = 0.$$

Hence, $f(x) = \sum_i a_i x^i$ is a nonzero polynomial over F having α as a zero. This shows that α is algebraic over F. We conclude that $E \supseteq F$ is an algebraic extension. □

12.6 Straightedge and compass constructions

In this section, we use the theory of field extensions to show that there cannot exist an algorithm for trisecting an arbitrary angle using only a straightedge and compass.

We need a careful description of what is meant by a straightedge and compass construction. Begin by choosing two arbitrary points on a piece of paper and call these two points O and I. Take these as the start of a collection of constructible points. New constructible points arise as intersections of lines and circles drawn with the straightedge and compass using points that have already been constructed. We call such lines and circles constructible. More precisely, a line is constructible if and only if it passes through two constructible points (intuitively, the two points are used to line up the straightedge) and a circle is constructible if and only if its center is a constructible point and it passes through a constructible point (intuitively, the point of the compass is placed on the first point and the compass is adjusted so that the drawing tip is at the second point).

We leave it to the reader to verify the following elementary fact: Given a constructible line L and a constructible point P, the line through P that is perpendicular to L is constructible, and so is the line through P parallel to L.

It is convenient to introduce a coordinate system. Turn the paper, if necessary, so that I is directly to the right of O. The line through these points is constructible; it forms the x-axis with O marking the origin and I marking the point $(1, 0)$. The line that passes through O and is perpendicular to the x-axis is constructible; it forms the y-axis. Finally, the circle with center O passing through I intersects the y-axis in the constructible point $J = (0, 1)$. This completes the setup of the coordinate system.

A real number α is **constructible** if it is a coordinate of a constructible point.

If a circle is constructible, then its radius is constructible: Given a constructible circle with center C, let P be the right point of intersection of the circle and a horizontal line through C. The line through P that is parallel to the line through O and C intersects the x-axis at the point $(r, 0)$, where r is the radius of the circle. (If C lies on the x-axis, then this construction fails, but the point $(0, r)$ can be constructed instead.)

12.6.1 Theorem. *If α is constructible, then $[\mathbf{Q}(\alpha) : \mathbf{Q}] = 2^n$ for some nonnegative integer n.*

Proof. Suppose that a certain sequence of constructions, as described above, have been carried out. Let F be a subfield of $\mathbf{R}$ that contains the coordinates of all currently constructed points and let (α, β) be a newly constructed point. We claim that the degree $[F(\alpha) : F]$ is either 1 or 2 (and likewise for $[F(\beta) : F]$).

First assume that (α, β) is obtained as a point of intersection of a constructible line $\mathcal{L}$ and a constructible circle $\mathcal{C}$. This line and this circle have equations

$$\mathcal{L}: \quad y - t_1 = \frac{t_2 - t_1}{s_2 - s_1}(x - s_1), \qquad \mathcal{C}: \quad (x - c_1)^2 + (y - c_2)^2 = r^2,$$

with $s_i, t_i, c_i, r \in F$ (unless the line is vertical, which case we leave to the reader to handle). Now $(x, y) = (\alpha, \beta)$ is a solution to both of theses equations, so, after substituting in and combining to eliminate β, one finds that α satisfies an equation of the form $a_0 + a_1\alpha + a_2\alpha^2 = 0$ with $a_i \in F$. Therefore,

α is a zero of $f(x) = a_0 + a_1 x + a_2 x^2 \in F[x]$. By Section 12.4, the minimal polynomial of α over F has degree 1 or 2, and so the degree $[F(\alpha) : F]$ is 1 or 2 as well. An analogous argument shows that the same is true of the degree $[F(\beta) : F]$.

In the two other cases, namely, (α, β) obtained as a point of intersection of two constructible lines or two constructible circles, similar reasoning leads to the same conclusion. Therefore, the claim is established.

We are now ready to prove the theorem. Let α be a constructible number. There is a finite sequence $(\alpha_1, \alpha_2), (\alpha_3, \alpha_4), \dots, (\alpha_{r-1}, \alpha_r)$ of constructible points such that the first point is constructed from O and I, each point after that is constructed from the preceding points, and either $\alpha_{r-1} = \alpha$ or $\alpha_r = \alpha$. Using the first part of the proof, we see that the degree of each field over the preceding field in the sequence

$$\mathbf{Q} \subseteq \mathbf{Q}(\alpha_1) \subseteq \mathbf{Q}(\alpha_1)(\alpha_2) \subseteq \cdots \subseteq \mathbf{Q}(\alpha_1)(\alpha_2)\cdots(\alpha_r) =: E$$

is either 1 or 2. Therefore $[E : \mathbf{Q}]$ is a power of 2 by Section 12.2. Since $\alpha \in E$, we have $E \supseteq \mathbf{Q}(\alpha) \supseteq \mathbf{Q}$, so, again by Section 12.2, $[\mathbf{Q}(\alpha) : \mathbf{Q}] = 2^n$ for some nonnegative integer n. □

12.6.2 Corollary. *There is no algorithm for trisecting an arbitrary angle using straightedge and compass.*

Proof. We argue by contradiction. Suppose that there is such an algorithm. Construct the equilateral triangle with base $\overline{OI}$, use the purported algorithm to trisect the 60° angle at O, and hence construct the point on the unit circle with x-coordinate $\cos 20^\circ$. Drop a perpendicular from this point to the x-axis to construct the point $(\cos 20^\circ, 0)$ and then use the circle through O having this point as center to construct the point $(2\cos 20^\circ, 0)$. Conclude that the number $\alpha = 2\cos 20^\circ$ is constructible.

From the trigonometric identity $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ we obtain

$$1/2 = \cos 60^\circ = 4\cos^3 20^\circ - 3\cos 20^\circ,$$

so that

$$1 = 8\cos^3 20^\circ - 6\cos 20^\circ = \alpha^3 - 3\alpha.$$

Therefore, α is a zero of $f(x) = x^3 - 3x - 1 \in \mathbf{Z}[x]$. If this polynomial is not irreducible over $\mathbf{Q}$, then it must have a linear factor and hence a zero in $\mathbf{Q}$, which is not the case since the rational root theorem (10.2) limits such zeros to ±1, and neither of these is a zero. Therefore, $f(x)$ is irreducible over $\mathbf{Q}$ and it follows that it is the minimal polynomial $p_\alpha(x)$ of α over $\mathbf{Q}$. By Section 12.4, $[\mathbf{Q}(\alpha):\mathbf{Q}]=3$, which is in conflict with the previous theorem.

We conclude that there is no algorithm for trisecting an angle using straightedge and compass. □

12 – Exercises

12–1 Let $E \supseteq F$ be a field extension, let $\alpha \in E$, and assume that the degree $[F(\alpha):F]$ is odd. Prove that $F(\alpha^2)=F(\alpha)$.

12–2 Prove that the set $\mathbf{A}=\{\alpha \in \mathbf{C} \mid \alpha \text{ is algebraic over } \mathbf{Q}\}$ is a subfield of $\mathbf{C}$.

Hint: For $\alpha, \beta \in \mathbf{A}$, first prove that the degree $[\mathbf{Q}(\alpha,\beta):\mathbf{Q}]$ is finite by using Section 12.2 with an appropriate intermediate field. Then use 12.5.

13 Galois correspondence

13.1 Definition: Galois group of extension

Let $E \supseteq F$ be fields. The set $\mathrm{Aut}(E)$ of all automorphisms of E is a group under function composition. If σ is an element of $\mathrm{Aut}(E)$ that fixes every element of F (meaning $\sigma(a)=a$ for all $a \in F$), then σ is an **F-automorphism** of E. An F-automorphism of E is bijective and it is simultaneously a ring homomorphism from E to itself and a linear transformation of the vector space E over F to itself.

The **Galois group** of E over F, denoted $\mathrm{Aut}_F(E)$, is the subgroup of $\mathrm{Aut}(E)$ consisting of all F-automorphisms of E. (We will sometimes refer to this group as the Galois group of the extension $E \supseteq F$.)

13.2 Examples

Here are some examples of Galois groups of extensions $E \supseteq F$.

- (E=F) In this case, the Galois group of the extension is the trivial group.

- ($\mathbf{C} \supseteq \mathbf{R}$) Let $\sigma \in \operatorname{Aut}_{\mathbf{R}}(\mathbf{C})$. We have $\sigma(i)^2 = \sigma(i^2) = \sigma(-1) = -1$, so $\sigma(i) = \pm i$. Therefore, for any $a + bi \in \mathbf{C}$,

  $$\sigma(a + bi) = \sigma(a) + \sigma(b)\sigma(i) = a \pm bi,$$

  implying that σ is either the identity map or complex conjugation. Conversely, the identity map and complex conjugation are both $\mathbf{R}$-automorphisms of $\mathbf{C}$, so the Galois group of $\mathbf{C}$ over $\mathbf{R}$ consists precisely of these two elements. We conclude that $\operatorname{Aut}_{\mathbf{R}}(\mathbf{C}) \cong \mathbf{Z}_2$.

- ($\mathbf{Q}(\sqrt{2}) \supseteq \mathbf{Q}$) By Section 12.4 the vector space $\mathbf{Q}(\sqrt{2})$ over $\mathbf{Q}$ has basis $\{1, \sqrt{2}\}$. Reasoning as in the previous example, we find that σ is in the Galois group of $\mathbf{Q}(\sqrt{2})$ over $\mathbf{Q}$ if and only if $\sigma(a+b\sqrt{2}) = a \pm b\sqrt{2}$, so this Galois group is also isomorphic to $\mathbf{Z}_2$.

- ($\mathbf{Q}(\sqrt[3]{2}) \supseteq \mathbf{Q}$) Put $\alpha = \sqrt[3]{2}$ and let $\sigma \in \operatorname{Aut}_{\mathbf{Q}}(\mathbf{Q}(\alpha))$. We have $\sigma(\alpha)^3 = \sigma(\alpha^3) = \sigma(2) = 2$, so $x = \sigma(\alpha)$ is a solution to the equation $x^3 = 2$. This solution is real since it lies in $\mathbf{Q}(\alpha) \subseteq \mathbf{R}$. But α is the only real solution to this equation, so $\sigma(\alpha) = \alpha$. Since every element of $\mathbf{Q}(\alpha)$ is a linear combination of powers of α with rational coefficients (12.4), it follows that σ is the identity map. We conclude that the Galois group of $\mathbf{Q}(\sqrt[3]{2})$ over $\mathbf{Q}$ is the trivial group.

13.3 Priming maps

Let $E \supseteq F$ be a field extension. A field M with $E \supseteq M \supseteq F$ is an **intermediate field** of the extension $E \supseteq F$. Let $\mathcal{M}$ be the set of all such intermediate fields. Put $G = \operatorname{Aut}_F(E)$ and let $\mathcal{H}$ be the set of all subgroups of G.

Define a map $\mathcal{M} \to \mathcal{H}$ by $M \mapsto M'$, where $M' = \operatorname{Aut}_M(E)$, the Galois group of E over M. Also, define a map in the reverse direction $\mathcal{H} \to \mathcal{M}$ by $H \mapsto H'$, where $H' = \{\alpha \in E \,|\, \sigma(\alpha) = \alpha \text{ for all } \sigma \in H\}$, the **fixed field** of the subgroup H. These are referred to as the **priming maps**; they can be visualized like this:

$$\begin{tikzcd}
E & \{\varepsilon\} & & E & \{\varepsilon\} \\
\cup & \cap & & \cup & \cap \\
M \arrow[r] & M' & & H' & H \arrow[l] \\
\cup & \cap & & \cup & \cap \\
F & G & & F & G
\end{tikzcd}$$

The priming maps are inclusion reversing, that is, if L and M are intermediate fields with $L \subseteq M$, then $L' \supseteq M'$, and similarly for the reverse map. The following result that relates intermediate field degrees to subgroup indexes is crucial for the development of the theory.

13.3.1 Theorem. *Assume that the extension $E \supseteq F$ is finite.*

(i) *If $M, L \in \mathcal{M}$ with $M \supseteq L$, then $|L' : M'| \leq [M : L]$.*

(ii) *If $H, J \in \mathcal{H}$ with $H \subseteq J$, then $[H' : J'] \leq |J : H|$.*

Proof. (i) Let $M, L \in \mathcal{M}$ with $M \supseteq L$. Assume that $M = L(\alpha)$ for some $\alpha \in M$. We will begin by proving the claim in this special case.

First, $[L(\alpha) : L]$ is finite by 12.2, so α is algebraic over L by 12.5.

Let $\sigma_1 M', \sigma_2 M', \ldots, \sigma_r M'$ be distinct left cosets of M' in L'. Let $1 \leq i, j \leq r$ and assume that $\sigma_i(\alpha) = \sigma_j(\alpha)$. Then $\sigma_j^{-1}\sigma_i$ fixes α and, since this automorphism is in L', it fixes all of $L(\alpha) = M$ (since every element of $L(\alpha)$ is a linear combination of powers of α with coefficients in L by 12.4). Therefore, $\sigma_j^{-1}\sigma_i \in M'$, implying that $\sigma_i M' = \sigma_j M'$, and so $i = j$. It follows that the elements $\sigma_1(\alpha), \sigma_2(\alpha), \ldots, \sigma_r(\alpha)$ are distinct.

Let $p_\alpha(x)$ be the minimal polynomial of α over L. By Exercise 13–1, $\sigma_i(\alpha)$ is a zero of $p_\alpha(x)$ for each i, so r is at most the number of zeros of $p_\alpha(x)$ in M, which is at most $\deg p_\alpha(x) = [L(\alpha) : L] = [M : L]$ (using 9.4 and 12.4). In summary, $r \leq [M : L]$.

We have shown that there can be at most $[M : L]$ distinct left cosets of M' in L', which implies $|L' : M'| \leq [M : L]$. This establishes the special case.

We now prove the claim by induction on $n = [M : L]$. If $n = 1$, then $M = L$ and the claim follows. Assume that $n > 1$. Then there exists $\alpha \in M$ with $\alpha \notin L$. If $M = L(\alpha)$, then the first part of the proof applies, so suppose that $M \neq L(\alpha)$. Then $[M : L(\alpha)]$ and $[L(\alpha) : L]$ are both strictly less than n (since neither is 1 and their product is n by 12.2), so the induction hypothesis gives

$$|L(\alpha)' : M'| \leq [M : L(\alpha)] \quad \text{and} \quad |L' : L(\alpha)'| \leq [L(\alpha) : L].$$

Therefore,

$$|L' : M'| = |L' : L(\alpha)'||L(\alpha)' : M'| \leq [L(\alpha) : L][M : L(\alpha)] = [M : L],$$

where we have used multiplicativity of subgroup index (from group theory).

(ii) Let $H, J \in \mathcal{H}$ with $H \subseteq J$. By part (i),

$$|\operatorname{Aut}_F(E)| = |\operatorname{Aut}_F(E) : \{\varepsilon\}| = |F' : E'| \leq |E : F| < \infty,$$

so the Galois group of the extension is finite. In particular, $|J : H|$ is finite.

Assume that the claim is false, that is, $[H' : J'] > |J : H| =: n$. Then there exist $\alpha_1, \alpha_2, \dots, \alpha_{n+1}$ in H' that are linearly independent over J'. Let $\sigma_1 H, \sigma_2 H, \dots \sigma_n H$ be the distinct left cosets of H in J with $\sigma_1 = \varepsilon$ (the identity map on E). The system

$$\begin{array}{ccccccccc}
\sigma_1(\alpha_1)c_1 & + & \sigma_1(\alpha_2)c_2 & + & \cdots & + & \sigma_1(\alpha_{n+1})c_{n+1} & = & 0 \\
\sigma_2(\alpha_1)c_1 & + & \sigma_2(\alpha_2)c_2 & + & \cdots & + & \sigma_2(\alpha_{n+1})c_{n+1} & = & 0 \\
\vdots & & \vdots & & & & \vdots & & \vdots \\
\sigma_n(\alpha_1)c_1 & + & \sigma_n(\alpha_2)c_2 & + & \cdots & + & \sigma_n(\alpha_{n+1})c_{n+1} & = & 0
\end{array}$$

has n equations and $n+1$ unknowns $c_i \in E$, so there exists a nontrivial solution. Among all such solutions, let $c_1, c_2, \dots, c_{n+1}$ be one with as many zeros as possible. We may (and do) make the following assumptions:

- there exists $1 \leq r \leq n+1$ such that for each i

$$c_i \neq 0 \text{ if } i \leq r \qquad \text{and} \qquad c_i = 0 \text{ if } i > r,$$

- $c_1 = 1$,
- $c_2 \notin J'$.

The first can be arranged for by relabeling the α_i, if necessary. Replacing each c_i by c_i/c_1 allows for the second. In view of our assumption that $\sigma_1 = \varepsilon$, the first equation of the system is $\sum_i c_i \alpha_i = 0$, so not all of the c_i lie in J'

(else linear independence of the α_i over J' is contradicted). Relabeling the α_i yet again, if necessary, allows for the third assumption.

There exists $\tau \in J$ such that $\tau(c_2) \neq c_2$. For each i, there exists a unique i' such that $\tau\sigma_i H = \sigma_{i'} H$. The map from $N := \{1, 2, \ldots, n\}$ to itself given by $i \mapsto i'$ is bijective. Indeed, for $i, j \in N$

$$\begin{aligned} i' = j' \;&\Rightarrow\; \sigma_{i'} H = \sigma_{j'} H \;\Rightarrow\; \tau\sigma_i H = \tau\sigma_j H \\ &\Rightarrow\; \sigma_i H = \sigma_j H \;\Rightarrow\; i = j, \end{aligned}$$

so that the map is injective and, since N is finite, the map is surjective as well.

Fix $i \in N$. We have $\tau\sigma_i = \sigma_{i'}\mu$ for some $\mu \in H$, so $\tau\sigma_i(\alpha_j) = \sigma_{i'}\mu(\alpha_j) = \sigma_{i'}(\alpha_j)$ for each j, where the last equality is due to the fact that $\alpha_j \in H'$. This establishes that $\tau\sigma_i(\alpha_j) = \sigma_{i'}(\alpha_j)$ for each i and each j.

Applying τ to the original system of equations yields

$$0 = \sum_j \tau\sigma_i(\alpha_j)\tau(c_j) = \sum_j \sigma_{i'}(\alpha_j)\tau(c_j)$$

($1 \leq i \leq n$), which shows that $\tau(c_1), \tau(c_2), \ldots, \tau(c_{n+1})$ is a solution to the system (since this is the same system with equations permuted). Subtracting from the earlier solution we obtain a solution

$$c_1 - \tau(c_1), c_2 - \tau(c_2), \ldots, c_{n+1} - \tau(c_{n+1}).$$

By our choices, $c_1 = 1$, so $c_1 - \tau(c_1) = 0$, and $\tau(c_2) \neq c_2$, so $c_2 - \tau(c_2) \neq 0$. It follows that this is a nontrivial solution to the system that has a greater number of zeros than the solution $c_1, c_2, \ldots, c_{n+1}$, contrary to our assumption. The proof is complete. □

13.4 Closed subfields and subgroups

Let $E \supseteq F$ be a field extension. It makes sense to compose the two priming maps of Section 13.3. In general, one has $M'' \supseteq M$ and $H'' \supseteq H$ for $M \in \mathcal{M}$ and $H \in \mathcal{H}$ (M'' is the set of field elements fixed by every automorphism fixing M; H'' is the set of automorphisms that fix every field element that is fixed by every automorphism in H).

An intermediate field $M \in \mathcal{M}$ is **closed** if $M'' = M$. Similarly, a subgroup $H \in \mathcal{H}$ is **closed** if $H'' = H$.

If all intermediate fields and subgroups were closed, then it would follow that the priming maps would be inverses of each other and would therefore define a one-to-one correspondence between intermediate fields and subgroups. Unfortunately, this is not the case in general as the following example shows.

- If $E = \mathbf{Q}(\sqrt[3]{2})$ and $F = \mathbf{Q}$, then F is not closed. Indeed, by Section 13.2, $F' = \mathrm{Aut}_F(E)$ is trivial, so $F'' = E \neq F$.

We can, nevertheless, impose conditions on the extension that will guarantee that all intermediate fields and all subgroups are closed. The extension $E \supseteq F$ is a **normal extension** if F is closed. By Exercise 13–3, the extension $E \supseteq F$ is normal if and only if each element of E that is not in F is moved by some element of the Galois group of the extension.

13.4.1 Theorem. *Assume that the extension $E \supseteq F$ is finite and normal.*

(i) *Every intermediate field $M \in \mathcal{M}$ and every subgroup $H \in \mathcal{H}$ is closed.*

(ii) *If $M, L \in \mathcal{M}$ with $M \supseteq L$, then $|L' : M'| = [M : L]$.*

(iii) *If $H, J \in \mathcal{H}$ with $H \subseteq J$, then $[H' : J'] = |J : H|$.*

Proof. Let $M \in \mathcal{M}$. Using Sections 12.2 and 13.3 we have $[M'' : F] < \infty$ and

$$[M'' : M][M : F] = [M'' : F] = [M'' : F''] \leq |F' : M'| \leq [M : F],$$

where we have used that $F'' = F$ since the extension $E \supseteq F$ is normal. This says that $[M'' : M] = 1$, which forces $M'' = M$. We conclude that every intermediate field of the extension is closed.

Let $M, L \in \mathcal{M}$ with $M \supseteq L$. Using what we have just proved, we have

$$[M : L] = [M'' : L''] \leq |L' : M'| \leq [M : L],$$

which forces $|L' : M'| = [M : L]$. This establishes the claims made about intermediate fields. The claims made about subgroups are proved similarly. □

13.5 Normality

Let $E \supseteq F$ be a finite, normal extension and put $G = \mathrm{Aut}_F(E)$. Let $M \in \mathcal{M}$ be an intermediate field of the extension.

13.5.1 Lemma. *The extension $M \supseteq F$ is normal if and only if $\sigma(M) \subseteq M$ for all $\sigma \in G$.*

Proof. Assume that $M \supseteq F$ is normal. Let $\alpha \in M$. Since the extension $E \supseteq F$ is finite, the element α is algebraic over F. We claim that the minimal polynomial $p_\alpha(x)$ of α over F factors as a product of linear factors over M.

For each $\tau \in H = \mathrm{Aut}_F(M)$, $\tau(\alpha)$ is a zero of $p_\alpha(x)$ (see Exercise 13–1). Therefore, the set $H(\alpha) = \{\tau(\alpha) \,|\, \tau \in H\}$ has at most $\deg p_\alpha(x)$ elements. Let $\alpha_1, \alpha_2, \dots, \alpha_m \in M$ be the distinct elements of this set. Then $m \leq \deg p_\alpha(x)$.

Put $f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_m)$. Let $\mu \in H$. Since $\mu(H(\alpha)) = (\mu H)(\alpha) = H(\alpha)$, we see that $\mu(\alpha_i) = \alpha_{i'}$, where $i \mapsto i'$ is some permutation of the set $\{1, 2, \dots, m\}$. With $\bar{\mu} : M[x] \to M[x]$ denoting the homomorphism induced by μ (see 9.8), we have

$$\begin{aligned} \mu f(x) = \bar{\mu}(f(x)) &= \bar{\mu}(\prod_i (x - \alpha_i)) = \prod_i \bar{\mu}(x - \alpha_i) \\ &= \prod_i (x - \mu(\alpha_i)) = \prod_i (x - \alpha_{i'}) = \prod_i (x - \alpha_i) = f(x), \end{aligned}$$

so that μ fixes the coefficients of $f(x)$. Since μ was arbitrary, we conclude that the coefficients of $f(x)$ are all contained in the fixed field of H, which is F'', where the priming operations are relative to the extension $M \supseteq F$. By our assumption that this extension is normal, we have $F'' = F$. Therefore, $f(x)$ is a polynomial over F. Now $\alpha_i = \alpha$ for some i (since $\alpha = \varepsilon(\alpha) \in H(\alpha)$), so α is a zero of $f(x)$. By Section 12.4, $p_\alpha(x)$ divides $f(x)$. However, $\deg f(x) = m \leq \deg p_\alpha(x)$ and $f(x)$ is monic, so $p_\alpha(x) = f(x) = \prod_i (x - \alpha_i)$ as claimed.

Let $\sigma \in G$. By Exercise 13–1, $\sigma(\alpha)$ is a zero of $p_\alpha(x)$. But we have just seen that $p_\alpha(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n)$ with $\alpha_i \in M$, so the α_i must be the only zeros of $p_\alpha(x)$ in E. Hence, $\sigma(\alpha) = \alpha_i \in M$ for some i. This shows that $\sigma(M) \subseteq M$ for each $\sigma \in G$.

Now assume that $\sigma(M) \subseteq M$ for each $\sigma \in G$. We prove that the extension $M \supseteq F$ is normal by using the characterization of normality stated in Exercise 13–3. Let α be an element of M that is not in F. There exists $\sigma \in G$ such that $\sigma(\alpha) \neq \alpha$ (by Exercise 13–3 and our assumption that the extension $E \supseteq F$ is normal). By our assumption, $\sigma(M) \subseteq M$. Since σ is an isomorphism of the vector space E over F onto itself, it follows that $\sigma(M)$ is a subspace of M isomorphic to M. Since M is finite dimensional over F, we conclude that $\sigma(M) = M$, that is, σ restricts to an element of $\mathrm{Aut}_F(M)$. By Exercise 13–3, the extension $M \supseteq F$ is normal. □

13.5.2 Theorem.

(i) *If H is a normal subgroup of G, then $H' \supseteq F$ is normal.*

(ii) *If $M \supseteq F$ is normal, then M' is a normal subgroup of G.*

Proof. (i) Let H be a normal subgroup of G. By the lemma, it suffices to show that $\sigma(H') \subseteq H'$ for each $\sigma \in G$. Let $\sigma \in G$ and let $\alpha \in H'$. For each $\mu \in H$, we have $\sigma^{-1}\mu\sigma \in H$ by normality of H, so that $\sigma^{-1}\mu\sigma(\alpha) = \alpha$, whence $\mu(\sigma(\alpha)) = \sigma(\alpha)$. Thus, $\sigma(\alpha) \in H'$ and the claim follows.

(ii) Assume that $M \supseteq F$ is normal. Let $\mu \in M'$ and $\sigma \in G$. For $\alpha \in M$, we have

$$\sigma^{-1}\mu\sigma(\alpha) = \sigma^{-1}(\mu(\sigma(\alpha))) = \sigma^{-1}(\sigma(\alpha)) = \alpha,$$

where we have used the lemma to see that $\sigma(\alpha)$ is in M and is therefore fixed by μ. Therefore, $\sigma^{-1}\mu\sigma \in M'$. This shows that M' is a normal subgroup of G as desired. □

13.6 Fundamental theorem of Galois theory

The following theorem is, for the most part, a summary of results stated and proved in earlier sections. The language here is somewhat informal; the reader desiring more precise statements can refer to the sections referenced in the proof.

Let $E \supseteq F$ be a finite, normal field extension and let $G = \mathrm{Aut}_F(E)$ be the Galois group of the extension.

13.6.1 Theorem (Fundamental theorem of Galois theory).

(i) *The priming maps define a one-to-one correspondence between the intermediate fields of the extension and the subgroups of G. This correspondence is inclusion reversing, and the degree of an intermediate extension equals the index of the corresponding subgroups. In particular, $|G| = [E : F]$.*

(ii) *If M is an intermediate field, then the extension $M \supseteq F$ is normal if and only if M' is a normal subgroup of G, and, in this case, $G/M' \cong \mathrm{Aut}_F(M)$.*

Proof. (i) That the priming maps define a one-to-one correspondence follows from the fact shown in 13.4 that all intermediate fields and all subgroups are closed since this implies that the priming maps are inverses of each other. The inclusion reversing property is immediate. The statement about degrees is established in 13.4. We have $[E : F] = |F' : E'| = |G : \{\varepsilon\}| = |G|$.

(ii) The fact that normal intermediate extensions correspond to normal subgroups is verified in 13.5.

All that remains is to check the stated isomorphism. Let M be an intermediate field and assume that $M \supseteq F$ is normal. Let $\varphi : G \to \mathrm{Aut}_F(M)$ be defined by $\varphi(\sigma) = \sigma|_M$ (restriction to M). In this definition, $\sigma \in G$ maps M into M by the lemma of Section 13.5, and the argument in the proof of that lemma shows that in fact $\sigma(M) = M$ so that $\sigma|_M \in \mathrm{Aut}_F(M)$. Therefore, φ is well defined. By the definition of M', we have $M' = \ker \varphi$. The first isomorphism theorem gives

$$G/M' = G/\ker\varphi \cong \operatorname{im}\varphi.$$

The proof is completed by showing that $\operatorname{im}\varphi = \mathrm{Aut}_F(M)$. Certainly, $\operatorname{im}\varphi$ is a subgroup of $\mathrm{Aut}_F(M)$. Using the isomorphism above and other parts of the theorem, we have $|\operatorname{im}\varphi| = |G : M'| = |F' : M'| = [M : F] = |\mathrm{Aut}_F(M)|$, where the last equality is from part (i) applied to the extension $M \supseteq F$. Therefore, $\operatorname{im}\varphi = \mathrm{Aut}_F(M)$ and the proof is complete. □

13.7 Example

Here, we compute the Galois group G of the extension $\mathbf{Q}(\alpha, \beta) \supseteq \mathbf{Q}$, where $\alpha = \sqrt{2}$ and $\beta = \sqrt{3}$.

We have a tower $\mathbf{Q}(\alpha, \beta) \supseteq \mathbf{Q}(\alpha) \supseteq \mathbf{Q}$. The minimal polynomial of α over $\mathbf{Q}$ is $x^2 - 2$, so $[\mathbf{Q}(\alpha) : \mathbf{Q}] = 2$. The minimal polynomial of β over $\mathbf{Q}(\alpha)$

is $x^2 - 3$. (Reason: Assume otherwise. Since β is a zero of $x^2 - 3$, the minimal polynomial must divide this polynomial and must therefore be the linear polynomial $x - \beta$, implying $\beta \in \mathbf{Q}(\alpha)$. Hence, $\beta = a + b\alpha$ for some $a, b \in \mathbf{Q}$. After ruling out the possibilities $a = 0$ or $b = 0$ we can square both sides of this equation and solve for α to get $\sqrt{2} = \alpha \in \mathbf{Q}$, a contradiction.) Therefore, $[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}(\alpha)] = 2$.

By the theorem in Section 12.2 and its proof, $[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}] = 4$ and the set $\{1, \alpha, \beta, \alpha\beta\}$ is a basis for $\mathbf{Q}(\alpha, \beta)$ over $\mathbf{Q}$. Using this basis, it is routine to show that there are elements σ and τ of the Galois group G satisfying

$$\begin{array}{rlcrl} \sigma : & \alpha \mapsto -\alpha & \qquad & \tau : & \alpha \mapsto \alpha \\ & \beta \mapsto \beta & & & \beta \mapsto -\beta. \end{array}$$

Let $\gamma \in \mathbf{Q}(\alpha, \beta)$. We have

$$\gamma = a_1 1 + a_2\alpha + a_3\beta + a_4\alpha\beta,$$

for some $a_i \in \mathbf{Q}$. If $\gamma \notin \mathbf{Q}$, then at least one of the coefficients a_2, a_3, or a_4 is nonzero, so that γ is not fixed either by σ or by τ. By Exercise 13–3, the extension $\mathbf{Q}(\alpha, \beta) \supseteq \mathbf{Q}$ is normal. Since the extension is finite as well the fundamental theorem of Galois theory (Section 13.6) applies.

We have $|G| = [\mathbf{Q}(\alpha, \beta) : \mathbf{Q}] = 4$. The automorphisms σ and τ each have order 2 and they are distinct. Since $|G| = 4$ it follows that σ and τ generate G and G is isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ with an isomorphism sending σ to $(1, 0)$ and τ to $(0, 1)$.

The Galois correspondence is represented here with primed objects occupying corresponding positions in the lattices:

$$\begin{tikzcd}[column sep=small] & \mathbf{Q}(\alpha, \beta) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\ \mathbf{Q}(\alpha) \arrow[dr, dash] & \mathbf{Q}(\alpha\beta) \arrow[d, dash] & \mathbf{Q}(\beta) \arrow[dl, dash] \\ & \mathbf{Q} & \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & \{\varepsilon\} \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\ \langle\tau\rangle \arrow[dr, dash] & \langle\tau\sigma\rangle \arrow[d, dash] & \langle\sigma\rangle \arrow[dl, dash] \\ & G & \end{tikzcd}$$

13 – Exercises

13–1 Let $E \supseteq F$ be a field extension, let $f(x)$ be a polynomial over F, let $\alpha \in E$ be a zero of $f(x)$, and let $\sigma \in \mathrm{Aut}_F(E)$. Prove that $\sigma(\alpha)$ is a zero of $f(x)$.

13–2 Prove that the Galois group of the extension $\mathbf{R} \supseteq \mathbf{Q}$ is the trivial group.

Hint: Prove, in this order, that an element of the Galois group (a) sends a square to a square, (b) sends a positive number to a positive number, and (c) preserves order. Then argue by contradiction using the fact that between any two real numbers there exists a rational number.

13–3 Let $E \supseteq F$ be a field extension and let $G = \mathrm{Aut}_F(E)$ be the Galois group of the extension. Prove that the extension is normal if and only if, for each element α of E that is not in F, there exists $\sigma \in G$ such that $\sigma(\alpha) \neq \alpha$.

14 Galois group of a polynomial

14.1 Fundamental theorem of algebra

From this point on, we will restrict our discussion to fields contained in the field $\mathbf{C}$ of complex numbers. Galois theory has been developed so far as to include more general fields, but we have chosen this restriction for the relative simplicity and ease of exposition that it offers. These advantages are due to the properties of the complex numbers stated in this section and the next.

The name given to the theorem below is a bit of a misnomer since there are no known proofs that involve only algebra. Nor does it seem likely that such a proof could exist. The complex numbers are defined in terms of the real numbers, which are defined using Cauchy sequences of rational numbers, so the notion of a limit is required. Therefore, at some point in every proof of the theorem (and there are many), analysis ultimately enters in.

The short proof given here is for the reader who has studied complex analysis.

14.1.1 Theorem (Fundamental theorem of algebra). *Every non-constant polynomial over the field* **C** *of complex numbers has a zero in* **C***.*

Proof. Let $f(x)$ be a polynomial over **C**. Assume that $f(x)$ has no zero in **C**. Then the complex-valued function $g(x) = 1/f(x)$ is defined. We have $\lim_{|x|\to\infty} |g(x)| = 0$, so $g(x)$ is a bounded entire function. According to Liouville's theorem, $g(x)$ is constant, say, $g(x) = c \in \mathbf{C}$. Therefore, $f(x) = 1/c$, a constant function. □

14.2 Irreducible polynomial has no multiple zeros in C

Let

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

be a polynomial of degree n over the field **C** of complex numbers.

14.2.1 Theorem.

(i) *The polynomial $f(x)$ factors as*

$$f(x) = a_n(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)$$

for some $\alpha_i \in$ **C***.*

(ii) *Let F be a subfield of* **C** *and assume that $f(x) \in F[x]$. If $f(x)$ is irreducible over F, then $f(x)$ has $\deg f(x)$ distinct zeros in* **C***.*

Proof. (i) The proof is by induction on n. There is nothing to prove if $n = 0$, so suppose that $n > 0$. By the fundamental theorem of algebra (14.1), $f(x)$ has a zero α_n in **C**. According to Section 9.4, $f(x) = g(x)(x - \alpha_n)$ for some polynomial $g(x)$ over **C**. By the induction hypothesis, $g(x) = a_n(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_{n-1})$ for some $\alpha_1, \alpha_2, \ldots, \alpha_{n-1} \in \mathbf{C}$. Therefore, $f(x)$ has the indicated factorization.

(ii) Assume that $f(x) = \sum_i a_ix^i$ is irreducible over F. In view of part (i) it is enough to show that $f(x)$ cannot be expressed in the form $f(x) = (x - \alpha)^2g(x)$ with $\alpha \in \mathbf{C}$ and $g(x)$ a polynomial over **C**. Suppose to the contrary that $f(x)$ has such an expression. In particular, α is a zero of $f(x)$, so that $\deg p_\alpha(x) = \deg f(x)$ (where $p_\alpha(x)$ is the minimal polynomial of α over F). The product rule for differentiation gives

$$f'(x) = 2(x-\alpha)g(x) + (x-\alpha)^2 g'(x),$$

which shows that α is a zero of $f'(x)$. Now $f'(x) = \sum_i ia_i x^{i-1}$, so $f'(x)$ is a polynomial over F. Moreover, the degree of $f'(x)$ is one less than that of $f(x)$, and $f'(x)$ is nonzero (since $f(x)$ is irreducible and hence nonconstant), so this contradicts the fact that the minimal polynomial $p_\alpha(x)$ is a nonzero polynomial over F of minimal degree having α as a zero. This contradiction completes the proof. □

14.3 Extending field isomorphisms

Let F and F_0 be subfields of the field of complex numbers and let $\sigma : F \to F_0$ be an isomorphism. Let $f(x)$ be an irreducible polynomial over F, let $\alpha \in \mathbf{C}$ be a zero of $f(x)$, and let β be a zero of $\sigma f(x)$, the polynomial obtained from $f(x)$ by applying σ to each coefficient (see 9.8).

14.3.1 Theorem. *There exists an isomorphism $\sigma_1 : F(\alpha) \to F_0(\beta)$ such that $\sigma_1|_F = \sigma$ and $\sigma_1(\alpha) = \beta$.*

Proof. The homomorphism $F[x] \to F_0[x]$ induced by σ is an isomorphism (since σ is) and it maps $f(x)$ to $\sigma f(x)$ and hence induces an isomorphism $F[x]/(f(x)) \to F_0[x]/(\sigma f(x))$. Using part (iv) of Section 12.4, as well as its proof, we have isomorphisms as indicated:

$$\begin{array}{ccccccc} F(\alpha) & \to & F[x]/(f(x)) & \to & F_0[x]/(\sigma f(x)) & \to & F_0(\beta), \\ \\ g(\alpha) & \mapsto & g(x) + (f(x)) & \mapsto & \sigma g(x) + (\sigma f(x)) & \mapsto & \sigma g(\beta). \end{array}$$

We have used the fact that $f(x)$ is an associate of $p_\alpha(x)$ so that $(p_\alpha(x)) = (f(x))$, and similarly $(p_\beta(x)) = (\sigma f(x))$.

Let $\sigma_1 : F(\alpha) \to F_0(\beta)$ be the composition of these isomorphisms. If $a \in F$, then letting $g(x)$ be the constant polynomial $g(x) = a$ shows that $\sigma_1(a) = \sigma(a)$, so that $\sigma_1|_F = \sigma$. Letting $g(x) = x$ shows that $\sigma_1(\alpha) = \beta$. □

14.4 Splitting field

Let F be a subfield of the field of complex numbers and let $f(x)$ be a nonconstant polynomial over F. The **splitting field** of $f(x)$ over F is the field

$$E = F(\alpha_1, \alpha_2, \dots, \alpha_n),$$

where the α_i are the zeros of $f(x)$ in $\mathbf{C}$. The terminology is due to the fact that over its splitting field the polynomial $f(x)$ "splits" into a product of linear factors (a consequence of Section 9.4) and this splitting field is the smallest subfield of $\mathbf{C}$ containing F over which $f(x)$ splits.

The extension $E \supseteq F$ is finite. Indeed, if for each $0 \leq i \leq n$ we put $M_i = F(\alpha_1, \alpha_2, \dots, \alpha_i)$, we get a tower

$$F = M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots \subseteq M_n = E$$

with each step $M_i \subseteq M_i(\alpha_{i+1}) = M_{i+1}$ finite, so Section 12.2 applies.

Let F_0 be another subfield of the field of complex numbers and let $\sigma : F \to F_0$ be an isomorphism. Let E be the splitting field of $f(x)$ over F and let E_0 be the splitting field of $\sigma f(x)$ over F_0.

14.4.1 Theorem. *There exists an isomorphism $\bar{\sigma} : E \to E_0$ such that $\bar{\sigma}|_F = \sigma$.*

Proof. The proof is by induction on $n = [E : F]$. Suppose $n = 1$. Then $E = F$, so that $f(x)$ splits into a product of linear factors over F. But then $\sigma f(x)$ splits into a product of linear factors over F_0, whence $E_0 = F_0$. Therefore, we can just let $\bar{\sigma} = \sigma$.

Assume that $n > 1$. Then $E \supsetneqq F$, so there exists a zero α of $f(x)$ that is in E but not in F. In the factorization of $f(x)$ as a product of irreducible polynomials over F (9.7), one of the factors, say $g(x)$, must have α as a zero. Now $\sigma g(x)$ is irreducible over F_0, so it is not constant and it therefore has a zero β in $\mathbf{C}$ (14.1). Since $\sigma g(x)$ is a factor $\sigma f(x)$, it follows that β is a zero of $\sigma f(x)$ as well, implying that $\beta \in E_0$.

By Section 14.3, there exists an isomorphism $\sigma_1 : F(\alpha) \to F_0(\beta)$ such that $\sigma_1|_F = \sigma$. It is immediate that E is the splitting field of $f(x)$ over $F(\alpha)$ and E_0 is the splitting field of $\sigma f(x)$ over $F_0(\beta)$. We have $[E : F] = [E : F(\alpha)][F(\alpha) : F]$ and $[F(\alpha) : F] > 1$, so $[E : F(\alpha)] < [E : F] = n$. By the induction hypothesis, there exists an isomorphism $\bar{\sigma} : E \to E_0$ such that $\bar{\sigma}|_{F(\alpha)} = \sigma_1$. Restricting both sides of this last equation to F gives $\bar{\sigma}|_F = \sigma$ as desired. □

14.5 Splitting field extension is normal

Let $E \supseteq F$ be an extension of fields contained in the field of complex numbers. The extension $E \supseteq F$ is a **splitting field extension** if E is the splitting field of some nonconstant polynomial over F.

As noted in the preceding section, such an extension is finite. The following theorem shows that it is normal as well. In particular, the fundamental theorem of Galois theory (13.6) applies to a splitting field extension.

14.5.1 Theorem. *If $E \supseteq F$ is a splitting field extension, then it is normal.*

Proof. Let $E \supseteq F$ be a splitting field extension, so that E is the splitting field of a nonconstant polynomial $f(x)$ over F. In order to show that this extension is normal, we need to show that $F'' = F$. In general, one has $F'' \supseteq F$, so there is a tower $E \supseteq F'' \supseteq F$ and hence a degree relationship $[E : F] = [E : F''][F'' : F]$. Therefore, it suffices to show that $[E : F''] = [E : F]$.

We observe that $F''' = F'$. Indeed, priming the relation $F'' \supseteq F$ and using the fact that a doubly primed subgroup contains the original produces $F''' \subseteq F' \subseteq F'''$, forcing equalities. Therefore, we have, using Section 13.3,

$$[E : F] \geq [E : F''] \geq |F''' : E'| = |F' : \{\varepsilon\}| = |G|,$$

where G is the Galois group of the extension. From this, we see that $[E : F''] = [E : F]$ would follow if we were to show that $|G| = [E : F]$. We claim that this latter is the case and proceed to give a proof by induction on $n = [E : F]$. If $n = 1$, then $E = F$ and $G = \{\varepsilon\}$, so both sides equal 1.

Assume that $n > 1$. Then $E \underset{\neq}{\supseteq} F$, so $f(x)$ has a zero α that is in E but not in F. As in the proof of 14.4, we find that $f(x)$ has an irreducible factor $g(x)$ over F having α as a zero. Moreover, $g(x)$ is not linear and it splits as a product of linear factors over E. By Section 14.2, $g(x)$ has $r = \deg g(x)$ (distinct) zeros $\alpha_1, \alpha_2, \dots, \alpha_r$ in E.

Put $M = F(\alpha)$ and $H = M'$. Let $\sigma_1 H, \sigma_2 H, \dots, \sigma_s H$ be the distinct cosets of H in G. Arguing as in the proof of part (ii) of 13.3, we have that $\sigma_1(\alpha), \sigma_2(\alpha), \dots, \sigma_s(\alpha)$ are distinct elements of E. By Exercise 13–1, each of these elements is a zero of $g(x)$ and hence equal to one of the elements α_i $(1 \leq i \leq r)$. This gives $s \leq r$. We will establish equality here by showing that each α_i equals $\sigma_j(\alpha)$ for some j.

Fix $1 \leq i \leq r$. By Section 14.3, with $\sigma : F \to F$ taken to be the identity map ε, we get an isomorphism $\tau : F(\alpha) \to F(\alpha_i)$ such that $\tau|_F = \sigma$ and $\tau(\alpha) = \alpha_i$. Now E is the splitting field of $f(x)$ over $F(\alpha)$, and E is also the splitting field of $f(x) = \tau f(x)$ over $F(\alpha_i)$, so Section 14.4 gives an isomorphism $\bar{\tau} : E \to E$ such that $\bar{\tau}|_{F(\alpha)} = \tau$. Restricting both sides of this last equation to F, we get $\bar{\tau}|_F = \sigma = \varepsilon$. Therefore, $\bar{\tau}$ is an F-automorphism of E, that is, $\bar{\tau} \in G$. This automorphism lies in the coset $\sigma_j H$ for some j, whence $\bar{\tau} = \sigma_j \mu$ for some $\mu \in H$. We have

$$\sigma_j(\alpha) = \sigma_j\mu(\alpha) = \bar{\tau}(\alpha) = \tau(\alpha) = \alpha_i.$$

In view of the preceding paragraph, we have $r = s$, so that

$$|G : H| = s = r = \deg g(x) = [F(\alpha) : F] = [M : F].$$

As noted above, E is the splitting field of $f(x)$ over M, and since $[E : M] < [E : M][M : F] = [E : F] = n$, the induction hypothesis applies to give $[E : M] = |H|$. Therefore,

$$|G| = |H||G : H| = [E : M][M : F] = [E : F],$$

and the proof is complete. □

14.6 Definition: Galois group of polynomial

Let F be a subfield of the field of complex numbers, let $f(x)$ be a nonconstant polynomial over F. The **Galois group** of $f(x)$ over F is the Galois group $G = \mathrm{Aut}_F(E)$ of the extension $E \supseteq F$, where E is the splitting field of $f(x)$ over F.

Assume that $f(x)$ is irreducible over F. By 14.2, $f(x)$ has $n = \deg f(x)$ (distinct) zeros $\alpha_1, \alpha_2, \ldots, \alpha_n$ in $\mathbf{C}$. Put $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$.

14.6.1 Theorem. *The function $\varphi : G \to \mathrm{Sym}(A)$ given by $\varphi(\sigma) = \sigma|_A$ is a group monomorphism.*

Proof. Let $\sigma \in G$. By Exercise 13–1, $\sigma(\alpha_i) \in A$ for each i, so σ maps A into A. Since σ is injective and A is finite, it follows that $\sigma|_A$ is a bijection $A \to A$, so $\sigma|_A \in \mathrm{Sym}(A)$ and φ is well defined.

For $\sigma, \tau \in G$, we have

$$\varphi(\sigma\tau) = (\sigma\tau)|_A = \sigma|_A\tau|_A = \varphi(\sigma)\varphi(\tau),$$

so φ is a homomorphism.

Let $\sigma \in \ker \varphi$. Then $\sigma|_A$ is the identity map on A, meaning that $\sigma(\alpha_i) = \alpha_i$ for each i. We prove, by induction on k, that $\sigma|_{E_k} = \varepsilon$ $(0 \leq k \leq n)$, where $E_k = F(\alpha_1, \alpha_2, \ldots, \alpha_k)$. The case $k = 0$ is immediate since $E_0 = F$.

Assume that $0 < k \leq n$. Let $\beta \in E_k$. Since $E_k = E_{k-1}(\alpha_k)$, β is a linear combination of powers of α_k with coefficients coming from E_{k-1} (see 12.4). By the induction hypothesis, σ fixes each of the coefficients in such a linear combination, and, since σ fixes α_k, it fixes powers of α_k as well. Therefore, $\sigma(\beta) = \beta$. This establishes the claim that $\sigma|_{E_k} = \varepsilon$. In particular, $\sigma = \sigma|_E = \sigma|_{E_n} = \varepsilon$. Thus, $\ker \varphi$ is trivial and φ is injective. The proof is complete. □

In practice, we use the monomorphism φ to identify the Galois group G of the irreducible polynomial $f(x)$ with a subgroup of the symmetric group $\mathrm{Sym}(A)$ and thereby view G as a group of permutations of the zeros of $f(x)$. By a further identification of α_i with i, we even view G as a subgroup of S_n.

14.7 Example

Let $f(x) = x^5 - 6x + 3 \in \mathbf{Q}[x]$. We claim that the Galois group G of $f(x)$ over $\mathbf{Q}$ is (isomorphic to) the symmetric group S_5.

First, $f(x)$ is irreducible over $\mathbf{Q}$ by Eisenstein's criterion with $p = 3$ (10.4). By Section 14.2, $f(x)$ has five distinct zeros $\alpha_1, \alpha_2, \ldots, \alpha_5$ in $\mathbf{C}$. Since $f(0) = 3$, $f(1) = -2$, and $f(2) = 23$, it follows that $f(x)$ has a real zero between 0 and 1 and a real zero between 1 and 2. Also, since $f''(x) = 20x^3$, the graph of $f(x)$ is concave down to the left of 0 and concave up to the right of 0, so $f(x)$ has a third real zero to the left of 0 and no other real zeros. We conclude that three of the α_i are real and the other two are nonreal.

Identify α_i with i and thereby view G as a subgroup of S_5 (see 14.6). Let E be the splitting field of $f(x)$ over $\mathbf{Q}$. We have $E = \mathbf{Q}(\alpha_1, \alpha_2, \ldots, \alpha_5) \supseteq \mathbf{Q}(\alpha_1)$, so, by the fundamental theorem of Galois theory (13.6) and Section 12.2,

$$|G| = [E : \mathbf{Q}] = [E : \mathbf{Q}(\alpha_1)][\mathbf{Q}(\alpha_1) : \mathbf{Q}].$$

It is evident that $f(x)$ is the minimal polynomial of α_1 over $\mathbf{Q}$, so the degree $[\mathbf{Q}(\alpha_1) : \mathbf{Q}]$ is 5 by Section 12.4. Therefore, 5 divides the order of G. By Cauchy's theorem (which says that if a prime number p divides the order of a group, then the group has an element of order p), G has an element of order 5, which must be a 5-cycle. By relabeling the α_i if necessary, we may (and do) assume that G contains the 5-cycle $(1, 2, 3, 4, 5)$ and that α_1 and α_2 are the two nonreal zeros of $f(x)$. The restriction of complex conjugation to E is an element of G that transposes α_1 and α_2 and leaves the real zeros fixed, so G also contains the transposition $(1, 2)$.

So far, we have that G contains the elements $\rho = (1, 2, 3, 4, 5)$ and $\tau = (1, 2)$. We claim that these two permutations generate S_5. Generally, if $\mu = (i_1, i_2, \dots, i_r)$ is a cycle in a symmetric group S_n and $\sigma \in S_n$, then the conjugate $\sigma\mu\sigma^{-1}$ equals $(\sigma(i_1), \sigma(i_2), \dots, \sigma(i_r))$. Using this observation one routinely checks that for each $1 \leq i < 5$, $(i, i+1) = \rho^{i-1}\tau\rho^{-(i-1)} \in G$. Again using the observation, it follows in turn that for each $1 \leq i < j \leq 5$, $(i, j) = \sigma(i, i+1)\sigma^{-1} \in G$, where $\sigma = (j-1, j)(j-2, j-1)\cdots(i+1, i+2)$. Since every element of S_5 is a product of transpositions, we conclude that $G = S_5$ as claimed.

14.8 Adjoining pth roots

Let F be a subfield of the field of complex numbers and let p be a prime number.

14.8.1 Theorem.

(i) *The Galois group of $x^p - 1$ over F is abelian.*

(ii) *If $a \in F$ and F contains the zeros of $x^p - 1$, then the Galois group of $x^p - a$ over F is abelian.*

Proof. (i) We have $x^p - 1 = (x - 1)(1 + x + x^2 + \cdots + x^{p-1})$ (cf. cyclotomic polynomial in 10.5). Since 1 is not a zero of the second factor, it follows that $x^p - 1$ has a zero ω in $\mathbf{C}$ with $\omega \neq 1$. Now $\omega^p = 1$, so the order of ω in the multiplicative group of nonzero complex numbers divides p. Since this order is not 1 and p is prime, we see that the order is precisely p. In particular, $1, \omega, \omega^2, \dots, \omega^{p-1}$ are distinct. For each i, $(\omega^i)^p = (\omega^p)^i = 1^i = 1$, implying that ω^i is a zero of the polynomial $x^p - 1$. We conclude that ω^i $(0 \leq i < p)$ are all of the zeros of $x^p - 1$ in $\mathbf{C}$. Therefore, the splitting field of $x^p - 1$ over F is $E := F(1, \omega, \omega^2, \dots, \omega^{p-1}) = F(\omega)$.

The Galois group of $x^p - 1$ over F is, by definition, the Galois group $G = \mathrm{Aut}_F(E)$ of the extension $E \supseteq F$. Let $\sigma, \tau \in G$. By Exercise 13–1, $\sigma(\omega) = \omega^i$

for some i. Similarly, $\tau(\omega) = \omega^j$ for some j. Therefore,

$$\begin{aligned}\sigma\tau(\omega) &= \sigma(\omega^j) = \sigma(\omega)^j = (\omega^i)^j \\ &= (\omega^j)^i = \tau(\omega)^i = \tau(\omega^i) = \tau\sigma(\omega).\end{aligned}$$

Since $E = F(\omega)$, an element of G is completely determined by its effect on ω. Thus, $\sigma\tau = \tau\sigma$. This shows that G is abelian as claimed.

(ii) Let $a \in F$ and assume that F contains the zeros of $x^p - 1$. Then $\omega^i \in F$ for each i, with ω as in the proof of part (i). Since $x^p - a$ is nonconstant, it has a zero α in $\mathbf{C}$. The numbers $\alpha\omega^i$ $(0 \leq i < p)$ are zeros of $x^p - a$ and they are distinct, so they are all of the zeros of $x^p - a$ in $\mathbf{C}$. It follows that the splitting field of $x^p - a$ over F is $E := F(\alpha, \alpha\omega, \alpha\omega^2, \dots, \alpha\omega^{p-1}) = F(\alpha)$.

Let $\sigma, \tau \in G$, where $G = \mathrm{Aut}_F(E)$, the Galois group of $x^p - a$ over F. We have $\sigma(\alpha) = \alpha\omega^i$ and $\tau(\alpha) = \alpha\omega^j$ for some i and j, so, using that the powers of ω are in F, we obtain

$$\begin{aligned}\sigma\tau(\alpha) &= \sigma(\alpha\omega^j) = \sigma(\alpha)\omega^j = \alpha\omega^i\omega^j \\ &= \alpha\omega^j\omega^i = \tau(\alpha)\omega^i = \tau(\alpha\omega^i) = \tau\sigma(\alpha).\end{aligned}$$

As in the proof of part (i), we conclude that G is abelian. □

15 Solvability by radicals

15.1 Motivation

The polynomial $f(x) = x^{10} - 2x^5 - 2$ over $\mathbf{Q}$ has $\sqrt[5]{1 + \sqrt{3}}$ as a zero. We imagine that this zero has been obtained from $\mathbf{Q}$ by using only extraction of roots $\sqrt[n]{}$ and field operations $+$, $-$, $\times$, $\div$ (actually, only $+$ in this case).

In order to make this process more precise, we proceed as follows: Let $\alpha_1 = \sqrt{3}$ and $\alpha_2 = \sqrt[5]{1 + \alpha_1}$ and note that the tower

$$\mathbf{Q} \subseteq \mathbf{Q}(\alpha_1) \subseteq \mathbf{Q}(\alpha_1, \alpha_2)$$

has the property that $\alpha_1^2 \in \mathbf{Q}$ and $\alpha_2^5 \in \mathbf{Q}(\alpha_1)$. In other words, the process can be thought of as an enlarging of the original field $\mathbf{Q}$ to the field $\mathbf{Q}(\alpha_1, \alpha_2)$ one step at a time, with each step being carried out by adjoining to the previous field a number having a power lying in that previous field, and with the final field containing the zero α_2 of the polynomial.

The extension $\mathbf{Q}(\alpha_1, \alpha_2) \supseteq \mathbf{Q}$ is an example of a radical extension (definition in 15.3). If all of the zeros of a given polynomial over a subfield F of the field of complex numbers lie in some radical extension of F, then it follows that all of these zeros can be obtained from F by using only extraction of roots and field operations, and we say that the polynomial is solvable by radicals over F.

There is a stringent condition on when a polynomial over F can be solvable by radicals over F, namely, the Galois group of the polynomial over F must be solvable (see 15.4). A solvable group (definition in 15.2) can be thought of as being built up of abelian groups. The proof of this condition is the main goal of this section.

An arbitrary quadratic polynomial $f(x) = ax^2+bx+c$ is solvable by radicals (as the quadratic formula shows), and the same is true of arbitrary cubic and quartic polynomials. However, it is not the case that every quintic polynomial is solvable by radicals. In other words, there does not exist an analog of the quadratic formula that will give all of the zeros of an arbitrary quintic polynomial. We prove this by exhibiting a particular quintic polynomial having a Galois group that is not solvable (see 15.5).

15.2 Solvable group

Let G be a group. A **solvable series** of G is a tuple $(G_i) = (G_0, G_1, \dots, G_r)$ satisfying

(i) $G_0 = G$,

(ii) $G_r = \{e\}$,

(iii) $G_i \triangleleft G_{i-1}$ for $0 < i \leq r$,

(iv) G_{i-1}/G_i is abelian for $0 < i \leq r$.

The group G is **solvable** if there exists a solvable series of G.

- If G is abelian, then it is solvable since (G_0, G_1) is a solvable series of G, where $G_0 = G$ and $G_1 = \{e\}$. This shows that the notion of solvable group generalizes the notion of abelian group.

- Let $G = S_3 = \{\varepsilon, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}$, the symmetric group of degree 3. Put

  $$G_0 = G, \quad G_1 = \langle (1,2,3) \rangle = \{\varepsilon, (1,2,3), (1,3,2)\}, \quad G_2 = \{\varepsilon\}.$$

  Since G_1 has index 2 in G, it is a normal subgroup. The quotient G_0/G_1 is isomorphic to $\mathbf{Z}_2$ and the quotient G_1/G_2 is isomorphic to $\mathbf{Z}_3$. Therefore, (G_0, G_1, G_2) is a solvable series of G and G is solvable.

- The alternating group A_5 (the subgroup of S_5 consisting of even permutations) is not solvable. In order to see this, we need the fact from group theory that A_5 is simple, that is, its only normal subgroups are the trivial subgroup and itself. Suppose that there exists a solvable series (G_i) of A_5. Then, for some i we must have $G_{i-1} = A_5$ and $G_i = \{\varepsilon\}$. This says that A_5 is abelian since it is isomorphic to G_{i-1}/G_i. However, $\sigma = (1,2,3)$ and $\tau = (3,4,5)$ are both even permutations (hence elements of A_5), and $\sigma\tau \neq \tau\sigma$, since $\sigma\tau(2) = 3$ and $\tau\sigma(2) = 4$, so A_5 is not abelian. This is a contradiction. Thus, there does not exist a solvable series of A_5 and A_5 is not solvable.

15.2.1 Theorem.

(i) *Let $H \leq G$. If G is solvable, then so is H.*

(ii) *Let $\varphi : G \to G'$ be a homomorphism. If G is solvable, then so is* $\operatorname{im} \varphi$.

(iii) *Let $N \triangleleft G$. If N and G/N are both solvable, then so is G.*

Proof. (i) Assume that G is solvable. There exists a solvable series (G_i) of G. For each i, put $H_i = H \cap G_i$. Fix i. Since $G_i \subseteq G_{i-1}$, we have $H \cap G_i = H \cap G_{i-1} \cap G_i = H_{i-1} \cap G_i$, and using the second isomorphism theorem for groups, we get

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

Since G_{i-1}/G_i is abelian, so is H_{i-1}/H_i. Therefore, (H_i) is a solvable series of H and H is solvable.

(ii) Assume that G is solvable. There exists a solvable series (G_i) of G. Fix i. We have an epimorphism $G_{i-1} \to \varphi(G_{i-1}) \to \varphi(G_{i-1})/\varphi(G_i)$, where the first map is the restriction of φ to G_{i-1} and the second is the canonical epimorphism. Since G_i is contained in the kernel of this map, the fundamental homomorphism theorem for groups provides an induced epimorphism $G_{i-1}/G_i \to \varphi(G_{i-1})/\varphi(G_i)$. Since G_{i-1}/G_i is abelian, so is $\varphi(G_{i-1})/\varphi(G_i)$. Therefore, $(\varphi(G_i))$ is a solvable series of $\varphi(G)$ and $\operatorname{im}\varphi = \varphi(G)$ is solvable.

(iii) Assume that N and G/N are both solvable. There exist solvable series (N_i) and (G_i') of N and G/N, respectively. Fix i. By the correspondence theorem for groups, $G_i' = G_i/N$ for some subgroup G_i of G containing N. By the third isomorphism theorem for groups, $G_{i-1}/G_i \cong (G_{i-1}/N)/(G_i/N) = G_{i-1}'/G_i'$. Therefore, $(G_0, G_1, \dots, N_0, N_1, \dots)$ is a solvable series of G and G is solvable. □

15.3 Radical extension

Let $M \supseteq F$ be an extension of fields contained in the field of complex numbers. The extension $M \supseteq F$ is a **radical extension** if there exists a tuple $(\alpha_1, \alpha_2, \dots, \alpha_r)$ of complex numbers such that

(i) $M = F(\alpha_1, \alpha_2, \dots, \alpha_r)$,

(ii) for each i, there exists a positive integer n_i such that

$$\alpha_i^{n_i} \in F(\alpha_1, \alpha_2, \dots, \alpha_{i-1}).$$

Assume that $M \supseteq F$ is a radical extension and let the notation be as above. Further, for each i, put $M_i = F(\alpha_1, \alpha_2, \dots, \alpha_i)$. Note that $M_i = M_{i-1}(\alpha_i)$ for each i.

We observe that $M \supseteq F$ is a finite extension. Indeed, for each i the element α_i is algebraic over M_{i-1} since it is a zero of the polynomial $x^{n_i} - \alpha_i^{n_i}$ over M_{i-1}, so that each step in the tower $M \supseteq M_{r-1} \supseteq M_{r-2} \supseteq \cdots \supseteq M_1 \supseteq F$ is finite.

15.3.1 Theorem. *If $E \supseteq F$ is a normal, radical extension, then its Galois group G is solvable.*

Proof. Let $E \supseteq F$ be a normal, radical extension. There exists a tuple $(\alpha_1, \alpha_2, \dots, \alpha_r)$ of complex numbers such that, with $E_i := F(\alpha_1, \alpha_2, \dots, \alpha_i)$ $(0 \leq i \leq r)$, we have $E = E_r$ and, for each $0 < i \leq r$ there exists a positive integer n_i such that $\alpha_i^{n_i} \in E_{i-1}$. Fix i and let $n_i = lm$ be a factorization of n_i. It is immediate that the augmented tuple $(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_i^l, \alpha_i, \dots, \alpha_r)$

continues to satisfy the properties in the definition of radical extension. Therefore, by inserting numbers in this fashion if necessary, we may (and do) assume that n_i is prime for each i.

We prove the theorem by induction on r. If $r = 0$, then $E = F$, so that G is trivial and hence solvable.

Assume that $r > 0$. We first assume that F contains the zeros of the polynomial $x^{n_1} - 1$. By the proof of (ii) in Section 14.8, the splitting field of the polynomial $x^{n_1} - \alpha_1^{n_1}$ over F equals $F(\alpha_1)$. The extension $F(\alpha_1) \supseteq F$ is normal (14.5), so, by the fundamental theorem of Galois theory (13.6), $F(\alpha_1)' = \mathrm{Aut}_{F(\alpha_1)}(E)$ is a normal subgroup of G and the corresponding quotient is isomorphic to the Galois group of $F(\alpha_1)$ over F, which is abelian (and hence solvable) by Section 14.8. In light of part (iii) of 15.2, we see that it suffices to show that the group $\mathrm{Aut}_{F(\alpha_1)}(E)$ is solvable. However, writing

$$E = F(\alpha_1, \alpha_2, \ldots, \alpha_r) = [F(\alpha_1)](\alpha_2, \ldots, \alpha_r)$$

reveals that $E \supseteq F(\alpha_1)$ is a radical extension. It is also a normal extension since the subfield $F(\alpha_1)$ of the field E is closed by the fundamental theorem of Galois theory (13.6). Since the tuple $(\alpha_2, \ldots, \alpha_r)$ has $r - 1$ elements, the induction hypothesis applies and the Galois group $\mathrm{Aut}_{F(\alpha_1)}(E)$ of this extension is solvable. This completes the proof of the special case where F contains the zeros of the polynomial $x^{n_1} - 1$.

Now we turn to the general case. By the proof of (i) in Section 14.8, the splitting field of the polynomial $x^{n_1} - 1$ over F equals $F(\omega)$ with ω a zero of $x^{n_1} - 1$.

We claim that the extension $E(\omega) \supseteq F$ is finite and normal. Each step in the tower $E(\omega) \supseteq E \supseteq F$ is finite, so $E(\omega) \supseteq F$ is finite. To check normality, we use the condition in Exercise 13–3. Let α be an element of $E(\omega)$ not in F. Since the extension $E \supseteq F$ is normal, there exists $\sigma \in G$ such that $\sigma(\alpha) \neq \alpha$. By Section 14.3, there exists an isomorphism $\bar{\sigma} : E(\omega) \to E(\omega)$ such that $\bar{\sigma}|_E = \sigma$. We have $\bar{\sigma} \in \mathrm{Aut}_F(E(\omega))$ and $\bar{\sigma}(\alpha) = \sigma(\alpha) \neq \alpha$. By the stated exercise, the extension $E(\omega) \supseteq F$ is normal.

Now, the extension $F(\omega) \supseteq F$ is normal (14.5), so, by the fundamental theorem of Galois theory (13.6), applied to the extension $E(\omega) \supseteq F$ (valid since this extension is finite and normal by the preceding paragraph), $F(\omega)'$ is a normal subgroup of $\mathrm{Aut}_F(E(\omega))$ and the corresponding quotient is isomorphic to the Galois group of $F(\omega)$ over F, which is abelian (and hence solvable) by Section 14.8. As in the special case, we conclude that it suffices to show that the group $\mathrm{Aut}_{F(\omega)}(E(\omega))$ is solvable. The extension $E(\omega) \supseteq F(\omega)$ is normal (13.6), and writing

$$E(\omega) = F(\alpha_1, \alpha_2, \ldots, \alpha_r)(\omega) = [F(\omega)](\alpha_1, \alpha_2, \ldots, \alpha_r),$$

we see that this extension is a radical extension as well. Since $F(\omega)$ contains the zeros of the polynomial $x^{n_1} - 1$, the special case proved above applies and therefore $\mathrm{Aut}_{F(\omega)}(E(\omega))$ is solvable. This completes the proof. □

15.4 Polynomial solvable by radicals has solvable Galois group

Let F be a subfield of the field of complex numbers and let $f(x)$ be a polynomial over F. The polynomial $f(x)$ is **solvable by radicals** over F if its splitting field is contained in a radical extension of F.

Our goal in this section is to show that, if $f(x)$ is solvable by radicals over F, then its Galois group over F is solvable. The main tool for the proof is the theorem in Section 15.3, which requires that the radical extension be normal. Although a radical extension need not be normal in general, as the extension $\mathbf{Q}(\sqrt[3]{2}) \supseteq \mathbf{Q}$ shows (see 13.4), the next lemma says that such an extension can at least be enlarged to a normal, radical extension, which is sufficient for our purposes.

15.4.1 Lemma. *Let $M \supseteq F$ be a radical extension. There exists a field extension $E \supseteq M$ such that $E \supseteq F$ is a normal, radical extension.*

Proof. Let the notation be as in the definition of radical extension, and for each i put $M_i = F(\alpha_1, \alpha_2, \ldots, \alpha_i)$. It was observed in Section 15.3 that $M \supseteq F$ is a finite extension. By Section 12.5, this extension is algebraic, so each α_i is algebraic over F.

Put $f(x) = p_{\alpha_1}(x)p_{\alpha_2}(x)\cdots p_{\alpha_r}(x)$, where $p_{\alpha_i}(x)$ is the minimal polynomial of α_i over F, and let $\beta_1, \beta_2, \ldots, \beta_n$ be the zeros of $f(x)$ in $\mathbf{C}$. Then $E := F(\beta_1, \beta_2, \ldots, \beta_n)$ is the splitting field of $f(x)$ over F. The extension $E \supseteq F$ is normal by Section 14.5. Moreover, for each i, the number α_i is a zero of $p_{\alpha_i}(x)$ and hence of $f(x)$, so $E \supseteq F(\alpha_1, \alpha_2, \ldots, \alpha_r) = M$.

Fix $1 \leq j \leq n$. The number β_j is a zero of $p_{\alpha_i}(x)$ for some i. By Section 14.3, the identity map $\varepsilon : F \to F$ extends to an isomorphism $F(\alpha_i) \to F(\beta_j)$ that sends α_i to β_j. By Section 14.4, this isomorphism extends to an isomorphism $\tau_j : E \to E$. By construction, $\tau_j(\alpha_i) = \beta_j$ and $\tau_j|_F = \varepsilon$.

We have

$$\begin{aligned} E = F(&\tau_1(\alpha_1), \tau_1(\alpha_2), \dots, \tau_1(\alpha_r), \\ &\tau_2(\alpha_1), \tau_2(\alpha_2), \dots, \tau_2(\alpha_r), \\ &\vdots \\ &\tau_n(\alpha_1), \tau_n(\alpha_2), \dots, \tau_n(\alpha_r)) \end{aligned}$$

since the right hand member is a subfield of E that contains F and β_j for each j and $E = F(\beta_1, \beta_2, \dots, \beta_n)$. Also, this equation shows that $E \supseteq F$ is a radical extension, since, for each j and i,

$$\tau_j(\alpha_i)^{n_i} = \tau_j(\alpha_i^{n_i}) \in \tau_j(M_{i-1}) \subseteq F(\tau_j(\alpha_1), \tau_j(\alpha_2), \dots, \tau_j(\alpha_{i-1})),$$

so $\tau_j(\alpha_i)^{n_i}$ is contained in the field obtained from F by adjoining all of the elements in the list that precede $\tau_j(\alpha_i)$. □

15.4.2 Theorem. *If $f(x)$ is solvable by radicals over F, then its Galois group over F is solvable.*

Proof. Assume that $f(x)$ is solvable by radicals over F. By definition, the splitting field L of $f(x)$ over F is contained in a radical extension M of F. By the lemma, there exists an extension E of M such that $E \supseteq F$ is a normal, radical extension. According to Section 15.3, the Galois group $G = \mathrm{Aut}_F(E)$ of the extension $E \supseteq F$ is solvable.

Now $L \supseteq F$ is normal (14.5), so, by the fundamental theorem of Galois theory, L' is a normal subgroup of G and the corresponding quotient G/L' is isomorphic to $\mathrm{Aut}_F(L)$, the Galois group of $f(x)$ over F. Finally, G/L' is the image of the solvable group G under the canonical epimorphism $\pi : G \to G/L'$, so $\mathrm{Aut}_F(L) \cong G/L'$ is solvable by part (ii) of Section 15.2. This completes the proof. □

15.5 Insolvability of the quintic

We end this section by exhibiting a quintic polynomial over $\mathbf{Q}$ that is not solvable by radicals over $\mathbf{Q}$.

Let $f(x) = x^5 - 6x + 3 \in \mathbf{Q}[x]$. By Section 14.7, the Galois group of $f(x)$ over $\mathbf{Q}$ is the symmetric group S_5. If S_5 were solvable, then its subgroup A_5 would be solvable as well (part (i) of 15.2), but this is not the case as was shown in 15.2. Therefore, the Galois group of $f(x)$ over $\mathbf{Q}$ is not solvable, which implies (15.4) that $f(x)$ is not solvable by radicals over $\mathbf{Q}$.

As pointed out in Section 15.1, this shows that there cannot exist an analog of the quadratic formula that gives the zeros of an arbitrary quintic polynomial over $\mathbf{Q}$. In fact, this is true of polynomials over $\mathbf{Q}$ of any degree greater than or equal to five, the proof using the fact that the alternating groups A_n are simple for $n \geq 5$, just as the proof of our special case used that A_5 is simple.

A Writing proofs

A.1 Strings of relations

In a string of relations, the main news value should appear at the ends of the string and all of the intermediate steps should be easily verifiable.

- *If* $r > 2$, *then* $r^2 + r - 6 = (r+3)(r-2) > 0$ $(r \in \mathbf{R})$.

  The point being made is that if r is greater than 2, then $r^2 + r - 6$ is positive. The equality $r^2+r-6 = (r+3)(r-2)$ is verified by multiplying out the right hand side; the inequality $(r+3)(r-2) > 0$ follows from the fact that both factors are positive under the assumption $r > 2$.

- $(2+3)^2 = 5^2 = 25 \neq 13 = 4 + 9 = 2^2 + 3^2$.

  This says that $(2+3)^2 \neq 2^2 + 3^2$.

- $\frac{1}{2} + \frac{2}{3} - \frac{1}{4} = \frac{6}{12} + \frac{8}{12} - \frac{3}{12} = \frac{11}{12} \notin \mathbf{Z}$.

  This says that $\frac{1}{2} + \frac{2}{3} - \frac{1}{4}$ is not an integer. It is confusing to the reader if this point is made by writing $\frac{11}{12} = \frac{6}{12} + \frac{8}{12} - \frac{3}{12} = \frac{1}{2} + \frac{2}{3} - \frac{1}{4} \notin \mathbf{Z}$. In working from left to right, he can easily check each step except for the last, $\frac{1}{2} + \frac{2}{3} - \frac{1}{4} \notin \mathbf{Z}$. For this, he has to work backwards to see that $\frac{1}{2} + \frac{2}{3} - \frac{1}{4}$ equals $\frac{11}{12}$ which is not an integer.

A.2 If *P*, then *Q*.

To prove a statement of the form "If P, then Q" (which is the same as "P implies Q"), assume that P is true and show that Q is true.

- *Let $a, b, c \in \mathbf{R}$. If $a < b$ and $c < 0$, then $ca > cb$.*

  Proof: Assume that $a < b$ and $c < 0$. Since $a < b$, we have $a - b < 0$. Therefore, $ca - cb = c(a - b) > 0$. Hence, $ca > cb$, as desired. □

A.3 P if and only if Q

A statement of the form "P if and only if Q" is a combination of the two statements "If P, then Q" and "If Q, then P," so it is often written with a double implication symbol: "$P \Leftrightarrow Q$." To prove such a statement, take each implication separately and proceed as in A.2.

- *For $r \in \mathbf{R}$, $r^2 - 2r = -1$ if and only if $r = 1$.*

  Proof: Let $r \in \mathbf{R}$.

  ($\Rightarrow$) Assume $r^2 - 2r = -1$. Then $(r - 1)^2 = r^2 - 2r + 1 = 0$, which implies $r - 1 = 0$. Hence, $r = 1$.

  ($\Leftarrow$) Assume $r = 1$. Then $r^2 - 2r = 1^2 - 2(1) = -1$. □

It is common to use ($\Rightarrow$) and ($\Leftarrow$) as above to introduce the particular implication being proved. Incidentally, you should convince yourself that ($\Leftarrow$) corresponds to the statement "P if Q" while ($\Rightarrow$) corresponds to the statement "P only if Q."

A.4 Counterexample

To show that a statement involving "for every" is false, provide a single, explicit counterexample.

- *For every positive real number r, we have $r^3 > r^2$.*

  This statement is false, for if $r = \frac{1}{2}$, then $r^3 = \frac{1}{8} \not> \frac{1}{4} = r^2$.

I could also have said that the statement is false, for if r is any real number less than 1, then $r^3 - r^2 = r^2(r - 1) < 0$, whence $r^3 < r^2$. However, the explicit counterexample above is preferable to this argument in that it is easier to understand and it says just what needs to be said.

A.5 Showing "there exists"

To prove a statement involving "there exists," just exhibit a single such object and show that it satisfies the stated property.

- *There exists an* $r \in \mathbf{R}$ *satisfying* $r^2 + r - 12 = 0$.

  Proof: Put $r = 3$. We have $r^2 + r - 12 = 3^2 + 3 - 12 = 0$.

Note that I did not tell the reader how I came up with an r that works. There is no obligation to reveal the thought process that leads to the insight. In fact, doing so risks confusing the reader since it is unexpected. Also, I did not include that $r = -4$ also works since exhibiting a single r sufficed.

A.6 Showing "for every"

To prove a statement involving "for every," start with an arbitrary such object and show that it satisfies the given property.

- *For every* $r \in \mathbf{R}$ *with* $r \geq 3$*, we have* $r^2 - 2r + 1 \geq 4$.

  Proof: Let $r \in \mathbf{R}$ with $r \geq 3$. Then $r^2 - 2r + 1 = (r-1)^2 \geq (3-1)^2 = 4$. □

The first sentence of the proof means "Let r denote an arbitrary (i.e., any old) real number greater than or equal to 3."

A.7 Proof by contradiction

There is a method for proving a statement called "Proof by contradiction" which is sometimes useful. To use this method, one assumes that the given statement is false and then proceeds to derive a contradiction. The contradiction signals the presence somewhere of an invalid step. Therefore, provided all the other steps are valid, one can conclude that the initial assumption was not correct, which is to say that the given statement is in fact true.

- *There are infinitely many prime numbers.* (A *prime number* is an integer greater than 1 that is evenly divisible by no positive integers except 1 and itself (e.g., 2, 3, 5, 7, 11, ...).)

Proof: Suppose the statement is false. In other words, suppose there are only finitely many primes. We may enumerate them: $p_1, p_2, \ldots, p_n$. Consider the number $s := p_1 p_2 \cdots p_n + 1$. Now s is an integer greater than 1, so it must be divisible by some prime, say p_i. This means that $s = p_i m$ for some integer m. But then, $1 = s - p_1 p_2 \cdots p_n = p_i(m - p_1 p_2 \cdots \hat{p_i} \cdots p_n)$ where the symbol $\hat{p_i}$ means "delete p_i." The expression in the parentheses is just some integer and, since it is not possible to multiply the prime p_i by another integer and get 1, this is an obvious contradiction. Hence, our original assumption is wrong, that is, there are infinitely many prime numbers. □

This is essentially Euclid's famous proof of the infinitude of primes.

A.8 Contrapositive

A statement of the form "If P, then Q" is logically equivalent to the statement "If not Q, then not P" meaning that the first statement is true if and only if the second statement is true (you should be able to convince yourself that this is the case). This second statement is called the *contrapositive* of the first. Sometimes, proving the contrapositive of a statement is easier than proving the statement itself.

- *If* $r \neq s$, *then* $2r + 3 \neq 2s + 3$ $(r, s \in \mathbf{R})$.

  Proof: We prove the contrapositive: If $2r + 3 = 2s + 3$, then $r = s$. Assume $2r + 3 = 2s + 3$. Subtracting 3 from both sides and dividing through by 2 gives $r = s$, as desired. □

Occasionally, people give a proof by contradiction (see A.7) of a statement that can be established more directly by proving its contrapositive. For example, to prove the above statement by contradiction, we would start off assuming that there exist $r, s \in \mathbf{R}$ such that $r \neq s$ and $2r + 3 = 2s + 3$. Then, as above, we would obtain $r = s$, contradicting that $r \neq s$. This proof is valid, but it is not as direct as the first proof. When a proof by contradiction ends up contradicting one of the initial assumptions, as in this case, it can usually be recast using the contrapositive. (Note that this was not the case in the example worked for A.7.)

A.9 Negation

In order to formulate the contrapositives of statements or to give proofs by contradiction, one needs to be able to negate statements. Usually, this is easy; for instance, the negative of $a = b$ is $a \neq b$. However, more complicated statements require some thought. Logicians have formal rules that can be used to accurately negate extremely complex statements, but since most statements occurring in mathematics have very simple logical structures, mathematicians tend not to use the formulas relying instead on their own reasoning. Statements involving "for every" sometimes cause problems, so here is an example.

- *$ab = ba$ for every $a, b \in G$.*

  The negative is "There exist $a, b \in G$ such that $ab \neq ba$" (not "$ab \neq ba$ for every $a, b \in G$").

A.10 Variable scope

The "scope" of a variable in a proof refers to the portion of the proof that starts where the variable is introduced and ends where the variable no longer has meaning.

Generally, if a variable x is introduced with "If $x\dots$" or "For every $x\dots$," then that variable (and every variable that depends on it), ceases to have meaning at the end of the sentence. Such a variable x is said to have "local scope."

On the other hand, a variable x introduced using "Let $x\dots$" or "There exists $x\dots$" has meaning all the way to the end of the proof. Such a variable is said to have "global scope."

- If n is an even integer, then $n = 2m$ for some integer m. Therefore, $m = n/2$.

  (Incorrect. Due to the conditional "If ..." the variable n has no meaning past the first sentence. Since m depends on this n, it too has no meaning past the first sentence.)

- Let n be an even integer. Then $n = 2m$ for some integer m. Therefore, $m = n/2$.

(Correct. The phrase "Let n be an even integer" fixes an arbitrary even integer, and from that point on n refers to that fixed even integer. The m in the next sentence is chosen to satisfy $n = 2m$, so it too continues to have meaning from that point on.)

- For every odd integer n, the integer $n+1$ is even. Therefore, $n+1 = 2m$ for some $m \in \mathbf{Z}$.

  (Incorrect. Due to the quantifier "For every," n ceases to have meaning past the first sentence.)

- Let n be an odd integer. Then $n + 1$ is even, so $n + 1 = 2m$ for some integer m. Therefore, $m = (n + 1)/2$.

  (Correct. Both n and m have the indicated meaning to the end of the proof, unless the meaning is overwritten by a new statement, such as "Let n be an even integer.")