STUDENT MATHEMATICAL LIBRARY
Volume 82

Problems in Abstract Algebra

A. R. Wadsworth

$$\mathbb{Z}[\sqrt{d}]/p\mathbb{Z}[\sqrt{d}] \cong \mathbb{Z}p[X]/(X^2 - [d]_p)$$

$$C_f = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} \end{pmatrix}$$

$$G(L/F) \cong G(K/F)/G(K/L)$$

AMS
American Mathematical Society

STUDENT MATHEMATICAL LIBRARY
Volume 82

Problems in Abstract Algebra

A. R. Wadsworth

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification.* Primary 00A07, 12-01, 13-01, 15-01, 20-01.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-82

Library of Congress Cataloging-in-Publication Data

Names: Wadsworth, Adrian R., 1947–
Title: Problems in abstract algebra / A. R. Wadsworth.
Description: Providence, Rhode Island: American Mathematical Society, [2017] | Series: Student mathematical library; volume 82 | Includes bibliographical references and index.
Identifiers: LCCN 2016057500 | ISBN 9781470435837 (alk. paper)
Subjects: LCSH: Algebra, Abstract – Textbooks. | AMS: General – General and miscellaneous specific topics – Problem books. msc | Field theory and polynomials – Instructional exposition (textbooks, tutorial papers, etc.). msc | Commutative algebra – Instructional exposition (textbooks, tutorial papers, etc.). msc | Linear and multilinear algebra; matrix theory – Instructional exposition (textbooks, tutorial papers, etc.). msc | Group theory and generalizations – Instructional exposition (textbooks, tutorial papers, etc.). msc
Classification: LCC QA162 .W33 2017 | DDC 512/.02–dc23 LC record available at https://lccn.loc.gov/2016057500

10 9 8 7 6 5 4 3 2 1 22 21 20 19 18 17

Contents

Preface

It is a truism that, for most students, solving problems is a vital part of learning a mathematical subject well. Furthermore, I think students learn the most from challenging problems that demand serious thought and help develop a deeper understanding of important ideas. When teaching abstract algebra, I found it frustrating that most textbooks did not have enough interesting or really demanding problems. This led me to provide regular supplementary problem handouts. The handouts usually included a few particularly challenging "optional problems," which the students were free to work on or not, but they would receive some extra credit for turning in correct solutions. My problem handouts were the primary source for the problems in this book. They were used in teaching Math 100, University of California, San Diego's yearlong honors level course sequence in abstract algebra, and for the first term of Math 200, the graduate abstract algebra sequence. I hope this problem book will be a useful resource for students learning abstract algebra and for professors wanting to go beyond their textbook's problems.

To make this book somewhat more self-contained and independent of any particular text, I have included definitions of most concepts and statements of key theorems (including a few short proofs). This will mitigate the problem that texts do not completely agree on some definitions; likewise, there are different names and versions of

some theorems. For example, what is here called the Fundamental Homomorphism Theorem many authors call the First Isomorphism Theorem; so what is here the First Isomorphism Theorem they call the Second Isomorphism Theorem, etc. References for omitted proofs are provided except when they can be found in any text.

Some of the problems given here appear as theorems or problems in many texts. They are included if they give results worth knowing or if they are building blocks for other problems. Many of the problems have multiple parts; this makes it possible to develop a topic more thoroughly than what can be done in just a few sentences. Multipart problems can also provide paths to more difficult results.

I would like to thank Richard Elman for helpful suggestions and references. I would also like to thank Skip Garibaldi for much valuable feedback and for suggesting some problems.

Introduction

This book is a collection of problems in abstract algebra for strong advanced undergraduates or beginning graduate students in mathematics. Some of the problems will be challenging even for very talented students. These problems can be used by students taking an abstract algebra course who want more challenge or some interesting enrichment to their course. They can also be used by more experienced students for review or to solidify their knowledge of the subject. Professors teaching algebra courses may use this book as a source to supplement the problems from their textbook.

The assumed background for those undertaking these problems includes familiarity with the basic set-theoretic language of mathematics and the ability to write rigorous mathematical proofs. For Chapters 4 and 5, rudimentary knowledge of linear algebra is also needed. Students should probably be taking or have taken an abstract algebra course or be reading an abstract algebra text concurrently.

No solutions are provided for the problems given here (though there are many hints). The guiding philosophy for this is that readers who do not succeed with a first effort at a difficult problem can often progress and learn more by going back to it at a later time. Solutions in the back of the book offer too much temptation to give up working on a problem too soon.

Here are some highlights of the problems below:

- Use of the Fibonacci sequence to estimate the efficiency of the Euclidean Algorithm; see problem 1.2.
- Analysis of generalized dihedral groups; see problems 2.12, 2.28, 2.68, 2.74.
- A number of problems on semidirect products of groups; see §2.7.
- A number of problems on presentations of groups by generators and relations; see §2.8 and problem 4.109.
- Determination of all nonabelian groups of order p^3, p prime; see problems 2.87–2.90 and 4.109.
- Exploration of the geometry of the quaternions, explaining why they are used in computer graphics; see problem 3.12.
- Construction of $\mathbb{R}$ from $\mathbb{Q}$ as the the ring of Cauchy sequences of rationals modulo the ideal of null sequences; in problems 3.23–3.25.
- Determination of the integers that are sums of two squares; see problem 3.62.
- Determination of all prime ideals of $\mathbb{Z}[X]$ (see problem 3.84) and of $\mathbb{Z}[\sqrt{-d}\,]$, for $d \in \mathbb{N}$; see problem 3.60.
- Use of factor spaces V/W in linear algebra throughout Chapter 4, which streamlines the proof of triangulability and diagonalizability results, as well as the Cayley–Hamilton Theorem and the derivation of the canonical forms.
- Exponentials of matrices, applied to the solution of systems of linear differential equations; see §4.10.
- A natural proof of the volume interpretation of determinants of real matrices, via the Singular Value Decomposition; see problem 4.108.
- Artin's Galois theoretic proof of the algebraic closure of $\mathbb{C}$; see problem 5.101.
- Proof of Quadratic Reciprocity using cyclotomic polynomials and discriminants; see problem 5.128.

- Characterization of constructible (by compass and straightedge) numbers by the Galois groups of their minimal polynomials over $\mathbb{Q}$; see problem 5.149.
- Proof of Hilbert's Nullstellensatz; see problem 5.37.

Topics not covered in these problems include modules over rings, chain conditions for rings and modules, integrality of commutative ring extensions, Dedekind domains, semisimplicity of noncommutative rings, categories, and homological algebra. For readers interested in pursuing any of these topics, see the Suggestions for Further Reading after Chapter 5 below.

0.1. Notation

Here is some standard notation that will be used throughout the book: For the basic number systems, let

$\mathbb{N} = \{1, 2, 3, \dots, \}$, the natural numbers (this is not entirely standard; many authors include 0 in $\mathbb{N}$, but we will not);

$\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$, the integers;

$\mathbb{Q} = \{k/n \mid k, n \in \mathbb{Z}, n \neq 0\}$, the rational numbers;

$\mathbb{R}$, the real numbers;

$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}, i^2 = -1\}$, the complex numbers.

We assume the reader is familiar with the basic language of sets and functions; also with cardinal numbers and the notion of countability; also with equivalence relations and their associated partitions. Here is some standard notation and terminology for sets and functions: Let A and B be sets. Then,

$A \cup B = \{c \mid c \in A \text{ or } c \in B\}$ is the union of A and B.

$A \cap B = \{c \mid c \in A \text{ and } c \in B\}$ is the intersection of A and B.

$A \times B = \{(a, b) \mid a \in A,\ b \in B\}$ is the Cartesian product of A and B.

$B \subseteq A$ means B is a subset of A.

$B \subsetneqq A$ means B is a proper subset of A, i.e., $B \subseteq A$ and $B \neq A$.

$B \supseteq A$ means $A \subseteq B$.

If $B \subseteq A$, then $A \setminus B = \{a \in A \mid a \notin B\}$ is the complement of B in A.

If $\{A_i\}_{i \in I}$ is a family of sets indexed by a set I, then $\prod_{i \in I} A_i$ is the Cartesian product of the A_i.

$\varnothing$ denotes the empty set.

$|A|$ denotes the cardinality of A, i.e., the number of elements of A. Usually, we will not distinguish between infinite cardinalities. Thus, either $|A|$ is a nonnegative integer or $|A| = \infty$.

Let $f\colon A \to B$ be a function. To define the function, we sometimes write $a \mapsto \ldots$, meaning that $f(a) = \ldots$ (where ... specifies some element of B). The *domain* of f is A. The *image* of f is

$$im(f) = \{f(a) \mid a \in A\}, \tag{0.1}$$

a subset of B. If $C \subseteq A$, then $f|_C$ denotes the *restriction* of f to C, which is the function $C \to B$ such that $f|_C(c) = f(c)$ for all $c \in C$. We say that f is *onto* (or *surjective*) if $im(f) = B$. We say that f is *one-to-one* (or *injective*) if $f(a_1) \neq f(a_2)$ whenever $a_1 \neq a_2$. We say that f is a *one-to-one correspondence* (or *bijective* function) if f is both one-to-one and onto. If $g\colon B \to C$ is another function, then the *composition of f and g* is the function

$$g \circ f\colon A \to C \quad \text{given by} \quad a \mapsto g(f(a)) \ \text{ for each } a \in A. \tag{0.2}$$

The *identity function* on A is

$$id_A\colon A \to A \quad \text{given by} \quad a \mapsto a \ \text{ for each } a \in A. \tag{0.3}$$

An *inverse function* for f is a function

$$f^{-1}\colon B \to A \ \text{ such that } \ f^{-1} \circ f = id_A \ \text{ and } \ f \circ f^{-1} = id_B. \tag{0.4}$$

Recall that f has an inverse function iff (i.e., if and only if) f is bijective, and there is then only one inverse function of f. Even if f is not bijective, for $b \in B$ and $D \subseteq B$, we write

$f^{-1}(b)$ for $\{a \in A \mid f(a) = b\}$, the *inverse image* of b in A;

$f^{-1}(D)$ for $\{a \in A \mid f(a) \in D\}$, the *inverse image* of D in A.

Here is a list of frequently used abbreviations:

FHT stands for Fundamental Homomorphism Theorem;

gcd stands for greatest common divisor;

IET	stands for	Isomorphism Extension Theorem;
iff	stands for	if and only if;
lcm	stands for	least common multiple;
PID	stands for	principal ideal domain;
resp.	stands for	respectively;
UFD	stands for	unique factorization domain.

0.2. Zorn's Lemma

For nearly all of the problems in this book, Zorn's Lemma is not needed. Nonetheless, there are a few significant (though inessential) results provable readily using Zorn's Lemma that are convenient to have available. These include: the existence of bases for infinite-dimensional vector spaces (see the end of this section); the existence of maximal ideals in nontrivial rings (see problem 3.43); and the existence and uniqueness up to isomorphism of the algebraic closure of a field (see problem 5.4 and Note 5.56). Therefore, we state Zorn's Lemma here, and will assume it as an axiom; a few problems below will require it.

In order to state Zorn's Lemma, we need some terminology for partially ordered sets: Let S be a set. A binary relation $\leq$ on S is called a *partial ordering* if for all $a, b, c \in S$

(i) $a \leq a$;

(ii) if $a \leq b$ and $b \leq c$, then $a \leq c$;

(iii) if $a \leq b$ and $b \leq a$, then $a = b$.

The partial ordering on S is said to be a *total ordering* if $a \leq b$ or $b \leq a$ for all $a, b \in S$. A *maximal element* of S is an $m \in S$ such that there is no $b \in S$ with $m \leq b$ and $m \neq b$. If T is a subset of S, an *upper bound* for T in S is a $u \in S$ with $t \leq u$ for each $t \in T$.

Here is *Zorn's Lemma*: Let S be a partially ordered set. If every totally ordered subset of S has an upper bound in S, then S has a maximal element.

We will take Zorn's Lemma as an axiom. This is reasonable, since it is known that Zorn's Lemma is equivalent to the Axiom of Choice

and to the Well Ordering Principle. (See, e.g., Kaplansky [**12**, pp. 58–64] or Rotman [**21**, Appendix] for proofs of these equivalences.) The *Axiom of Choice* asserts that if $\{A_i\}_{i\in I}$ is a collection of nonempty sets indexed by a set I, then their Cartesian product $\prod_{i\in I} A_i$ is also nonempty. Since the Axiom of Choice is very reasonable (indeed, it seems intuitively evident) and is known not to introduce any contradictions in set theory, it is natural to accept Zorn's Lemma.

For an example of a typical application of Zorn's Lemma we now sketch a proof that every vector space has a base. (See the beginning of Chapter 4 if you are unfamiliar with the terminology used here.) Let V be a vector space over a field F. Let S be the set of all linearly independent subsets of V, partially ordered by inclusion. That is, for $A, B \in S$, we set $A \leq B$ just when $A \subseteq B$. If $\{A_i\}_{i\in I}$ is a totally ordered subset of S, Let $U = \bigcup_{i\in I} A_i$. Recall that U is linearly independent iff every finite subset U_0 of U is linearly independent. But every such U_0 lies in some A_j, since the collection of A_i is totally ordered; then U_0 is linearly independent, as A_j is linearly independent. Thus, U is linearly independent, which implies that $U \in S$. Clearly, U is an upper bound for $\{A_i\}_{i\in I}$. Hence, Zorn's Lemma implies that S has a maximal element, call it $\mathcal{B}$. Suppose there were $v \in V$ with v not a linear combination of the elements of $\mathcal{B}$. Then, as $\mathcal{B}$ is linearly independent, $\mathcal{B} \cup \{v\}$ is also linearly independent. Since $\mathcal{B} \subsetneqq \mathcal{B} \cup \{v\}$, this contradicts the maximality of $\mathcal{B}$ in S. Hence, there is no such v, i.e., $\mathcal{B}$ spans V. Since $\mathcal{B}$ is also linearly independent, it is a base of V.

Chapter 1

Integers and Integers mod n

The *Fibonacci sequence* is the infinite sequence of integers f_i for $i = 1, 2, \ldots$ defined recursively by

$$f_1 = 1, \;\; f_2 = 1, \;\; \text{and} \;\; f_n \;=\; f_{n-2} + f_{n-1} \;\; \text{for all integers } n \geq 2. \quad (1.1)$$

Thus, the initial terms in the Fibonacci sequence are:

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ \ldots.$$

The integer f_n is called the *n-th Fibonacci number*.

1.1. Prove the closed formula for the Fibonacci sequence:

$$f_n \;=\; \tfrac{1}{\sqrt{5}}\left(\tfrac{1+\sqrt{5}}{2}\right)^n - \tfrac{1}{\sqrt{5}}\left(\tfrac{1-\sqrt{5}}{2}\right)^n \quad (1.2)$$

(i) Prove formula (1.2) by mathematical induction.

(ii) Prove formula (1.2) by using its generating function: This is the formal power series

$$g \;=\; \sum_{i=1}^{\infty} f_i X^i \;=\; X + X^2 + 2X^3 + 3X^4 + 5X^5 + 8X^6 + \ldots .$$

Use the recursive formula for the f_i to express g as a quotient of polynomials. Then find a new series expansion for g using its partial fractions decomposition.

For another proof of the closed formula (1.2) for the Fibonacci sequence, using linear algebra, see problem 4.44 below. Note that since $\left|(1-\sqrt{5}\,)/2\right| < 1$ and $\frac{1}{\sqrt{5}}\left|\frac{1-\sqrt{5}}{2}\right| < 0.5$ it follows that f_n is the integer nearest $\frac{1}{\sqrt{5}}\big((1+\sqrt{5}\,)/2\big)^n$. For example,

$$f_{1000} \approx \tfrac{1}{\sqrt{5}}\big((1+\sqrt{5}\,)/2\big)^{1000} \approx 4.34666 \times 10^{208},$$

so f_{1000} is a 209-digit number.

The *Division Algorithm* for integers says that for any given integers a, b with $b \geq 1$ there exist unique integers q and r such that

$$a = qb + r, \quad \text{with} \quad 0 \leq r \leq b-1. \tag{1.3}$$

Recall that for integers a and b, we say that a *divides* b (denoted $a|b$) if there is an integer c with $b = ca$. When a and b are nonzero, the *greatest common divisor* of a and b (denoted $gcd(a,b)$) is the largest positive integer dividing both a and b. The *least common multiple* of a and b (denoted $lcm(a,b)$) is the smallest positive integer that is a multiple of both a and b.

Recall the *Euclidean Algorithm* for computing greatest common divisors by repeated application of the Division Algorithm: Take any nonzero integers a, b with $b \geq 1$ (and without loss of generality, $b \leq |a|$). By the Division Algorithm, we can write successively

$$\begin{aligned}
a &= q_1 b + r_1 \quad \text{with} \quad 0 < r_1 \leq b-1;\\
b &= q_2 r_1 + r_2 \quad \text{with} \quad 0 < r_2 \leq r_1 - 1;\\
r_1 &= q_3 r_2 + r_3 \quad \text{with} \quad 0 < r_3 \leq r_2 - 1;\\
&\cdots\\
r_{j-2} &= q_j r_{j-1} + r_j \quad \text{with} \quad 0 < r_j \leq r_{j-1} - 1;\\
&\cdots\\
r_{n-3} &= q_{n-1} r_{n-2} + r_{n-1} \quad \text{with} \quad 0 < r_{n-1} \leq r_{n-2} - 1;\\
r_{n-2} &= q_n r_{n-1} + 0.
\end{aligned}$$

The repeated division process terminates when the remainder r_n hits 0. The process must terminate after finitely many steps because $b > r_1 > r_2 > \ldots \geq 0$. Then,

$$gcd(a,b) = r_{n-1}, \text{ the last nonzero remainder.}$$

The Euclidean Algorithm shows the existence of $gcd(a, b)$ and also that $gcd(a, b)$ is expressible as $sa + tb$ for some integers s, t. The number n of times the Division Algorithm is applied is called the *number of steps* needed in computing $gcd(a, b)$.

For example, let $f_1, f_2, \ldots$ be the Fibonacci sequence. For an integer $i \geq 2$, the number of steps needed in computing that

$$gcd(f_i, f_{i+1}) = 1$$

is $i - 1$. (The successive remainders r_j in the long divisions are $f_{i-1}, f_{i-2}, \ldots, f_3, f_2, 0$.)

1.2. *Efficiency of the Euclidean Algorithm.* Take any nonzero integers a, b with $1 \leq b \leq |a|$, and let n be the number of steps needed in computing $gcd(a, b)$, as defined above. Let f_j be the j-th Fibonacci number. Prove that if $b \leq f_j$, for $j \geq 2$, then $n \leq j - 1$.

The preceding example shows that the bound on n in problem 1.2 is the best possible. This problem shows that determination of greatest common divisors is very efficient from a computational standpoint. Recall that f_n is the integer nearest $\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$. So, if b is a k-digit number, then the number of divisions required to compute $gcd(a, b)$ is at most

$$\big(k + log_{10}(\sqrt{5}\,)\big)\big/log_{10}\big((1 + \sqrt{5}\,)/2\big) \;\approx\; 4.785k + 1.672.$$

For example, if b is 100-digit number, then $gcd(a, b)$ can be computed with at most 481 long divisions.

1.3. Let m, n be positive integers with $gcd(m, n) = 1$. Determine the least integer k such that every integer $\ell \geq k$ is expressible as $\ell = rm + sn$ for some *nonnegative* integers r, s.

For example, if $m = 5$ and $n = 8$, then $k = 28$. Thus, with a supply of 5-cent and 8-cent stamps, one can make exact postage for any amount of 28 cents or more, but not for 27 cents.

Congruence mod n. Fix a positive integer n. For $a, b \in \mathbb{Z}$ we say that a and b are *congruent modulo* n, denoted

$$a \equiv b \;(mod\; n), \tag{1.4}$$

if $n|(b - a)$, i.e., there is some $t \in \mathbb{Z}$ with $b - a = tn$.

Recall the *Chinese Remainder Theorem*, which says that for all $m, n \in \mathbb{N}$ with $gcd(m, n) = 1$ and any $a, b \in \mathbb{Z}$ the there is an $x \in \mathbb{Z}$ such that

$$x \equiv a \ (mod\ m) \quad \text{and} \quad x \equiv b \ (mod\ n). \tag{1.5}$$

Moreover, any $x' \in \mathbb{Z}$ satisfies the same congruence conditions as x in (1.5) iff $x' \equiv x \ (mod\ mn)$. (See Example 2.19 below for a proof of the Chinese Remainder Theorem.)

1.4. Take any $m, n \in \mathbb{N}$ and let $d = gcd(m, n)$. Prove that for any $a, b \in \mathbb{Z}$ there is an $x \in \mathbb{Z}$ with $x \equiv a \ (mod\ m)$ and $x \equiv b \ (mod\ n)$ iff $a \equiv b \ (mod\ d)$. Moreover, when this holds, any $x' \in \mathbb{Z}$ satisfies the same congruence conditions as x iff $x' \equiv x \ (mod\ \ lcm(a, b))$.

$\mathbb{Z}_n$. Fix $n \in \mathbb{N}$. For $a \in \mathbb{Z}$ let $[a]_n$ denote the *congruence class of* $a \ (mod\ n)$, i.e.,

$$[a]_n \ = \ \{c \in \mathbb{Z} \mid c \equiv a \ (mod\ n)\} \ = \ \{a + tn \mid t \in \mathbb{Z}\}. \tag{1.6}$$

Since $[a]_n$ is the equivalence class of a with respect to the equivalence relation on $\mathbb{Z}$ given by congruence $(mod\ n)$, we have

$$[a]_n \ = \ [b]_n \quad \text{iff} \quad a \ \equiv \ b \ (mod\ n), \tag{1.7}$$

and the congruence classes $(mod\ n)$ form a partition of $\mathbb{Z}$. Let

$$\mathbb{Z}_n \ = \{[a]_n \mid a \in \mathbb{Z}\}. \tag{1.8}$$

The Division Algorithm shows that

$$\mathbb{Z}_n \ = \ \{[0]_n, [1]_n, \ldots, [n-1]_n\}$$

and that the classes $[0]_n, [1]_n, \ldots, [n-1]_n$ are distinct. Thus, $\left|\mathbb{Z}_n\right| = n$. There are operations of addition and multiplication on $\mathbb{Z}_n$ given by

$$[a]_n + [b]_n \ = \ [a + b]_n \quad \text{and} \quad [a]_n \cdot [b]_n \ = \ [ab]_n, \tag{1.9}$$

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

1.5. *Well-definition of Z_n operations.*

(i) The formula for $[a]_n + [b]_n$ in (1.9) is expressed in terms of a and b. But the choice of the integer a to describe $[a]_n$ is not unique (see (1.7)). That the sum is *well-defined* means that if $[a]_n = [a']_n$ and $[b]_n = [b']_n$, then we get the same congruence class for the sum whether the sum is determined

using a and b or a' and b', i.e., $[a+b]_n = [a'+b']_n$. Prove this.

(ii) Prove that the product operation given in (1.9) is well-defined.

1.6. Fix $n \in \mathbb{N}$ and take any $k \in \mathbb{Z}$ with $gcd(k,n) = 1$. Prove that for any congruence class $[a]_n$ in $\mathbb{Z}_n$ there is a unique congruence class $[b]_n$ such that $[k]_n \cdot [b]_n = [a]_n$.

1.7.

(i) Prove Wilson's Theorem: If p is a prime number, then

$$(p-1)! \equiv -1 \ (mod\ p).$$

(ii) Prove that if $n \in \mathbb{N}$ is not a prime number, then

$$(n-1)! \equiv 0 \ (mod\ n).$$

Euler's φ-function (also called Euler's totient function) is the map $\varphi\colon \mathbb{N} \to \mathbb{N}$ given by

$$\varphi(n) \ = \ \big|\{k \in \mathbb{N} \mid 1 \le k \le n \text{ and } gcd(k,n) = 1\}\big|. \tag{1.10}$$

Note that for any prime number p and any $r \in \mathbb{N}$, $\varphi(p) = p-1$ and $\varphi(p^r) = p^r - p^{r-1}$. Since $\varphi(mn) = \varphi(m)\varphi(n)$ whenever $gcd(m,n) = 1$ (see problem 1.9 below), it follows that for distinct prime numbers $p_1, \ldots, p_k$ and positive integers $r_1, \ldots, r_k$ if $n = p_1^{r_1} \ldots p_k^{r_k}$, then

$$\varphi(n) \ = \ \prod_{j=1}^{k} \big(p_j^{r_j} - p_j^{r_j-1}\big) \ = \ n \cdot \big(1 - \tfrac{1}{p_1}\big) \ldots \big(1 - \tfrac{1}{p_k}\big). \tag{1.11}$$

1.8. Take any $m, n \in \mathbb{N}$ with $gcd(m,n) = 1$. Prove that for $k \in \mathbb{Z}$, $gcd(k,mn) = 1$ iff $gcd(k,m) = 1$ and $gcd(k,n) = 1$.

1.9. Prove that for $m, n \in \mathbb{N}$,

$$\text{if } gcd(m,n) = 1, \quad \text{then} \quad \varphi(mn) \ = \ \varphi(m)\varphi(n). \tag{1.12}$$

(For a proof of this formula using groups, see (2.16) below.)

1.10. Prove that for any $n \in \mathbb{N}$,

$$\sum_{d|n} \varphi(d) \ = \ n. \tag{1.13}$$

The sum is taken over all the divisors d of n with $1 \le d \le n$. (See Example 2.18 below for a group-theoretic approach to this formula.)

Recall that for integers n, k with $0 \leq k \leq n$ the *binomial coefficient* $\binom{n}{k}$ is defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k(k-1)(k-2)\ldots 1}. \tag{1.14}$$

An easy calculation from the definition yields *Pascal's Identity*:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \qquad \text{for all } \ n, k \in \mathbb{N} \text{ with } 1 \leq k < n. \tag{1.15}$$

It follows by induction on n that $\binom{n}{k}$ is always an integer. Note that

$$\text{if } p \text{ is a prime number and } 1 \leq k \leq p-1, \text{ then } \ p \mid \binom{p}{k}. \tag{1.16}$$

For $p \mid p\binom{p-1}{k} = (p-k)\binom{p}{k}$. Since p is prime and $p \nmid (p-k)$, p must divide $\binom{p}{k}$.

1.11. Prove *Fermat's Theorem*: If p is a prime number, then

$$a^p \equiv a \ (mod\ p) \qquad \text{for any} \quad a \in \mathbb{Z}.$$

(Hint: Prove this by induction on a using the binomial expansion.)

See problem 2.4(ii) below for another proof of Fermat's Theorem.

Chapter 2

Groups

This chapter has problems on groups. In a few places, rudimentary facts about rings and fields are needed. These are all recalled in Chapter 3 below.

2.1. Groups, subgroups, and cosets

Recall some basic terminology and facts about groups and subgroups. A *group* is a nonempty set G with a binary operation, denoted $\cdot$, such that

(i) the operation is *associative*, i.e.,

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \qquad \text{for all } a, b, c \in G;$$

(ii) there is an *identity element* 1_G of G satisfying

$$1_G \cdot a = a \cdot 1_G = a \qquad \text{for every } a \in G;$$

(iii) every $a \in G$ has an *inverse* $b \in G$ satisfying $a \cdot b = b \cdot a = 1_G$. This b is uniquely determined by a and is denoted a^{-1}.

The group G is said to be *abelian* if $a \cdot b = b \cdot a$ for all $a, b \in G$. The *order* of G is the number of elements of G, and is denoted $|G|$. Then either $|G| \in \mathbb{N}$ or $|G| = \infty$.

For any a, b in a group G we write ab for $a \cdot b$ when the group operation is clear, and we often write 1 for 1_G. Also, for $a, b, c \in G$

we write abc for $(ab)c$ which equals $a(bc)$, and inductively for $n \geq 4$, $a_1a_2 \dots a_n$ denotes $(a_1a_2 \dots a_{n-1})a_n$. For $a \in G$, and $k \in \mathbb{Z}$ we define a^k by: $a^0 = 1$, $a^1 = a$ and if $k > 1$ inductively $a^k = a^{k-1}a$, and if $k < 0$, $a^k = (a^{-k})^{-1}$. The following usual laws of exponents hold: for all $j, k \in \mathbb{Z}$, $a^{j+k} = a^j a^k$ and $a^{jk} = (a^j)^k$. However, $(ab)^k = a^k b^k$ holds for all $k \in \mathbb{Z}$ only when $ab = ba$. For $a \in G$, the *order* of a, denoted $|a|$, is the least positive integer n such that $a^n = 1$, if there is such an n. If there is no such n, we set $|a| = \infty$. Recall that

$$\text{if } |a| = \infty \text{ then } a^j = a^k \text{ iff } j = k, \text{ for all } j, k \in \mathbb{Z}.$$

But,

$$\text{if } a \text{ has finite order } n \text{ then } a^j = a^k \text{ iff } j \equiv k \ (mod\ n).$$

Subgroups and cosets. A nonempty subset H of a group G is called a *subgroup* of G if for any $h, h' \in H$, $hh' \in H$ and $h^{-1} \in H$. When this occurs, H is a group with the operation of G, and $1_H = 1_G \in H$. When $H \subsetneqq G$, H is called a *proper subgroup* of G. The *trivial subgroup* of G is $\{1_G\}$. Now fix a subgroup H of G. For any $g \in G$, the set

$$gH = \{gh \mid h \in H\} \tag{2.1}$$

is called the *left coset* of H in G determined by g. (Likewise, the right coset of H determined by g is $Hg = \{hg \mid h \in H\}$.) The left cosets of H in G form a partition of G, since they are the equivalence classes for the equivalence relation $\sim$ on G given by $g \sim g'$ just when $g^{-1}g' \in H$. (Then gH is the equivalence class of g.) The left cosets of H in G are in general not the same as the right cosets, but the well-defined map $gH \mapsto Hg^{-1}$ gives a one-to-one correspondence between the left cosets and the right cosets. The *index* of H in G, denoted $|G{:}H|$, is the number of left cosets of H in G (= the number of right cosets of H in G). Since $|gH| = |H|$for each $g \in G$, we have *Lagrange's Theorem*, which says that

$$|G| = |G{:}H| \cdot |H|. \tag{2.2}$$

Additive notation. There are some groups, (e.g., $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or any vector space over a field), where the group operation is customarily written as addition. When the operation on a group G is denoted $+$, we write 0_G for the identity element of G and $-a$ for the inverse of $a \in G$, and write ka instead of a^k for $k \in \mathbb{Z}$. Also, $a + b$

is never abbreviated as ab. We say $(G, +)$ is a group to indicate that the operation is written additively. Additive notation is used only for abelian groups. If H is a subgroup of G, we write

$$a + H = \{a + h \mid h \in H\}$$

for the left (= right) coset of H in G determined by a.

A *normal subgroup* of G is a subgroup N such that $gN = Ng$ for every $g \in G$. Equivalently, subgroup N is normal in G if $gNg^{-1} = N$ for every $g \in G$, where $gNg^{-1} = \{gng^{-1} \mid n \in N\}$.

For any groups G and H, a function $f\colon G \to H$ is a *group homomorphism* if $f(g_1 \cdot g_2) = f(g_1) \cdot f(g_2)$ for all $g_1, g_2 \in G$. When this occurs, $f(1_G) = 1_H$ and $f(g^{-1}) = f(g)^{-1}$ for all $g \in G$. If the homomorphism f is bijective (= one-to-one and onto), we say that f is a group *isomorphism*. When there is an isomorphism $G \to H$ we say that G and H are *isomorphic*, and write

$$G \cong H.$$

For more on homomorphisms and normal subgroups, see §2.2.

Here is a list of some of the standard examples of groups. Throughout the list, n is any positive integer.

- $\Sigma(S)$ denotes the *symmetric group* of a set S, which is the set of bijective (= one-to-one and onto) functions $f\colon S \to S$. The group operation is composition of functions. (Since we always write functions on the left, for $f, g \in \Sigma(S)$, the composition $f \circ g$ is given by $s \mapsto f(g(s))$.) The identity element of $\Sigma(S)$ is the identity map id_S. The inverse of $f \in S$ is its inverse function f^{-1}.
- $S_n = \Sigma(\{1, 2, 3, \ldots, n\})$is the *$n$-th symmetric group*.
- $(R, +)$ denotes the additive group of a ring R (e.g., $R = \mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$, or $R = M_n(T)$ the ring of $n \times n$ matrices over a ring T).
- $(V, +)$ denotes the additive group of a vector space V over any field F.

- $(\mathbb{Z}_n, +)$ is a group, where the operation is given by
$$[i]_n + [j]_n = [i + j]_n.$$
- R^* denotes the (*multiplicative*) *group of units* of a ring R, i.e.,
$$R^* = \{\, r \in R \mid \text{there is } s \in R \text{ such that } rs = sr = 1_R \,\}.$$
The group operation is the ring multiplication, and the identity element is 1_R. Thus, if F is a field such as $\mathbb{R}$ or $\mathbb{C}$ or $\mathbb{Q}$, then $F^* = F \setminus \{0\}$. Also, $\mathbb{Z}^* = \{1, -1\}$.
- $GL_n(R)$ denotes the *general linear group of R of degree n*, which is the group of invertible $n \times n$ matrices over R, for any ring R. A matrix $A \in M_n(R)$ is *invertible* over R if there is a matrix $B \in M_n(R)$ such that $AB = BA = I_n$, where I_n is the identity matrix in $M_n(R)$. Note that $GL_n(R) = M_n(R)^*$.
- $SL_n(R)$ denotes the *special linear group of degree n of R*, for any commutative ring R. This is the subgroup of $GL_n(R)$ consisting of matrices of determinant 1_R.
- O_n denotes the *orthogonal group of degree n over the real numbers*. This is $\{A \in M_n(\mathbb{R}) \mid A^t A = I_n\}$, where A^t is the transpose of the matrix A. The group operation on O_n is matrix multiplication, and O_n is a subgroup of $GL_n(\mathbb{R})$.
- $\langle S \rangle$ for a subset S of a group G, denotes *the subgroup of G generated by S*, i.e., $\langle S \rangle$ consists of 1_G and all the products of elements of S and their inverses.
- C_n denotes the *cyclic group of order n*, i.e., for
$$\omega_n = e^{2\pi i/n} = cos(2\pi/n) + i\, sin(2\pi/n) \in \mathbb{C},$$
C_n is the subgroup of $\mathbb{C}^*$ of order n generated by ω_n.

Example 2.1. *Subgroups of* $\mathbb{Z}$. The integers $\mathbb{Z}$ are an abelian group with respect to addition, For $n \in \mathbb{Z}$, let
$$n\mathbb{Z} = \{nr \mid r \in \mathbb{Z}\},$$
which is the subgroup of $\mathbb{Z}$ generated by n. Since $(-n)\mathbb{Z} = n\mathbb{Z}$, it suffices to restrict attention to $n \geq 0$. Note that $0\mathbb{Z} = \{0\}$, the trivial subgroup of $\mathbb{Z}$, and the cosets of $0\mathbb{Z}$ in $\mathbb{Z}$ are the singleton

sets: $k + 0\mathbb{Z} = \{k\}$ for $k \in \mathbb{Z}$. For $n > 0$, the cosets of $n\mathbb{Z}$ are the congruence classes *mod* n, as in (1.6)

$$k + n\mathbb{Z} = \{k + rn \mid r \in \mathbb{Z}\} = [k]_n.$$

Since there are n congruence classes *mod* n, $|\mathbb{Z}{:}n\mathbb{Z}| = n$ for $n > 0$. Hence, for distinct positive integers m and n, $m\mathbb{Z} \neq n\mathbb{Z}$. Note that for $m, n \in \mathbb{N}$

$$\langle m, n\rangle = \{rm + sn \mid r, s \in \mathbb{Z}\} = d\mathbb{Z}, \quad \text{where } d = gcd(m, n).$$

It follows from this that the groups $k\mathbb{Z}$ for $k \geq 0$ are *all* the subgroups of $\mathbb{Z}$. (Alternatively, if H is a nontrivial subgroup of $\mathbb{Z}$, and n is the least positive element of H, then one can use the Division Algorithm to verify that $H = n\mathbb{Z}$.) Thus, every subgroup of $\mathbb{Z}$ is *cyclic*, i.e., generated by a single element. Also, for $m, n \in \mathbb{N}$,

$$m\mathbb{Z} \cap n\mathbb{Z} = \ell\mathbb{Z}, \quad \text{where } \ell = lcm(m, n).$$

Note that $m\mathbb{Z} \subseteq n\mathbb{Z}$ iff $n|m$.

2.2. Let G be a group, and let $a \in G$.

(i) Prove that if $|a| = \infty$, then $\left|a^k\right| = \infty$ for each nonzero $k \in \mathbb{Z}$.

(ii) Prove that if $|a| = n < \infty$, then for each nonzero $k \in \mathbb{Z}$

$$\left|a^k\right| = n/gcd(k, n) = lcm(k, n)/|k|.$$

In particular, $\langle a^k\rangle = \langle a\rangle$ iff $\left|a^k\right| = n$ iff $gcd(k, n) = 1$. Thus, the number of elements of $\langle a\rangle$ which generate $\langle a\rangle$ is $\varphi(n)$, where φ is Euler's φ-function as in (1.10).

2.3. $\mathbb{Z}_n^*$. For any $n \in \mathbb{N}$, there is a well-defined associative operation on $\mathbb{Z}_n$ of multiplication given by $[i]_n \cdot [j]_n = [ij]_n$ for all $i, j \in \mathbb{Z}$ (recall problem 1.5(ii)), and $[1]_n$ is an identity element for this operation. However, $\mathbb{Z}_n$ is not a group with this operation, since not every element has an inverse. To get a group we must restrict to those elements that do have multiplicative inverses: The subset of $\mathbb{Z}_n$

$$\mathbb{Z}_n^* = \{[k]_n \mid \text{there is } [\ell]_n \in \mathbb{Z}_n \text{ with } [k]_n \cdot [\ell]_n = [1]_n\} \tag{2.3}$$

is easily seen to be an abelian group with respect to multiplication. Prove that

$$\mathbb{Z}_n^* = \{\, [k]_n \in \mathbb{Z}_n \mid gcd(k, n) = 1\}; \tag{2.4}$$

hence,

$$\left|\mathbb{Z}_n^*\right| = \varphi(n), \tag{2.5}$$

where φ is Euler's φ-function.

2.4. Take any integer $n \geq 2$.

(i) Prove *Euler's Theorem*: For any $k \in \mathbb{Z}$ if $gcd(k, n) = 1$, then $k^{\varphi(n)} \equiv 1 \ (mod\ n)$, where φ is Euler's φ-function.

(ii) Deduce *Fermat's Theorem*: If p is any prime number, and $a \in \mathbb{Z}$, then $a^p \equiv a \ (mod\ p)$ (cf. problem 1.11).

2.5. *Public key encryption.* Let p and q be distinct prime numbers, and let $r = pq$; so, $\varphi(r) = (p-1)(q-1)$. Choose any $c \in \mathbb{N}$ with $gcd(c, \varphi(r)) = 1$, and choose any $d \in \mathbb{N}$ with $cd \equiv 1 \ (mod\ \varphi(r))$. Prove that for any $a \in \mathbb{N}$,

$$a^{cd} \equiv a \ (mod\ r).$$

This formula is the basis of the RSA (for Rivest, Shamir, and Adelman) public key encryption system, which is widely used: Think of $a \in \mathbb{N}$ as a message to be securely transmitted, with $a < r$. Let $b = a^c$, and let b' be the remainder on dividing b by r; so $b' \equiv b \ (mod\, r)$ and $0 \leq b' < r$. This b' is the encoded message. The encoded message is decoded by someone knowing r and d by computing b'^d and taking the remainder on dividing by r, which yields the message a. The integers $r = pq$ and c and the encoded message b' can be public information. But to do the decoding one must know d as well (along with r). To be able to break the code knowing r and c, one would have to know $\varphi(r)$ to be able to determine $d \ (mod\ \varphi(r))$. For this it would suffice to determine the primes p and q from $r = pq$. But this is extremely difficult when the primes p and q are large (in practice on the order of 100 digits or more) since there is no computationally efficient method of computing prime factorizations.

2.6. Let H and K be subgroups of a group G, and let

$$HK = \{hk \mid h \in H,\ k \in K\}.$$

Prove that HK is a subgroup of G iff $HK = KH$. (This holds, in particular, if H or K is a normal subgroup of G.)

2.7. Let G be a group, and let H and K be subgroups of G.

(i) Prove that for any left cosets aH of H and bK of K for $a, b \in G$, either $aH \cap bK = \varnothing$ or $aH \cap bK$ is a coset of $H \cap K$.

(ii) Prove that if $|G{:}H| < \infty$ and $|G{:}K| < \infty$, then

$$|G : H \cap K| \le |G{:}H|\,|G{:}K| < \infty.$$

(iii) For any $a, b \in G$ prove that $b^{-1}Kb$ is a subgroup of G and that either $aH \cap Kb = \varnothing$ or $aH \cap Kb$ is a left coset of $H \cap b^{-1}Kb$.

2.8. Let G be an infinite group, and let $H_1, \ldots, H_n$ be subgroups of G (not necessarily distinct). Suppose G is a union of cosets, one for each H_i, say $G = a_1H_1 \cup a_2H_2 \cup \ldots \cup a_nH_n$.

(i) Prove that there is an i with $|G{:}H_i| < \infty$.

(ii) Prove that each coset a_jH_j with $|G{:}H_j| = \infty$ is redundant in the union, i.e., G is the union of the other cosets besides this one.

(iii) Prove that for some i, $|G{:}H_i| \le n$.

For any group G, the *center of* G is

$$Z(G) = \{g \in G \mid ag = ga \text{ for every } a \in G\}, \tag{2.6}$$

which is clearly an abelian normal subgroup of G.

2.9. *Generalized quaternion groups.* Take any integer $n \ge 2$, and let $\omega = e^{\pi i/n} = cos(\frac{\pi}{n}) + isin(\frac{\pi}{n})$, an element of order $2n$ in $\mathbb{C}^*$. In $GL_2(\mathbb{C})$, let

$$c = \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \quad \text{and} \quad d = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note that $c^{2n} = I_2$ (the identity matrix), $d^2 = c^n$, and $dcd^{-1} = c^{-1}$. Let

$$Q_n = \langle c, d \rangle,$$

a subgroup of $GL_2(\mathbb{C})$. This Q_n is called a *generalized quaternion group*. (The group Q_2 is isomorphic to the usual *quaternion group*, the subgroup $\{\pm 1, \pm \mathbf{i}, \pm \mathbf{j}, \pm \mathbf{k}\}$ of $\mathbb{H}^*$, see problem 3.11 below.)

(i) Prove that $\left|Q_n\right| = 4n$.

(ii) Prove that every element of $Q_n \setminus \langle c \rangle$ has order 4.

(iii) Prove that $\langle d^2 \rangle$ is the only subgroup of Q_n of order 2, and that $\langle d^2 \rangle = Z(Q_n)$.

(iv) Prove that if n is a 2-power, then every nontrivial subgroup of Q_n contains $\langle d^2 \rangle$.

(v) Prove that every subgroup of the nonabelian group Q_2 is a normal subgroup.

2.10. The (real) *orthogonal group of degree* 2 is

$$O_2 \;=\; \{A \in M_2(\mathbb{R}) \mid AA^t = I_2\},$$

where A^t is the transpose of the matrix A and I_2 is the identity matrix in $M_2(\mathbb{R})$.

(i) Prove that for any $A \in O_2$,

$$A \;=\; \begin{pmatrix} c & -\varepsilon s \\ s & \varepsilon c \end{pmatrix} \tag{2.7}$$

for some $c, s \in \mathbb{R}$ with $c^2 + s^2 = 1$ and $\varepsilon = \pm 1$.

(ii) Left multiplication by A defines an $\mathbb{R}$-linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$, where $\mathbb{R}^2 = \left\{ \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \,\middle|\, x, y \in \mathbb{R} \right\}$. Prove that when $\varepsilon = 1$ in (2.7) this transformation is rotation about the origin $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ counterclockwise through an angle θ, where $c = \mathit{cos}(\theta)$ and $s = \mathit{sin}(\theta)$ with $\theta > 0$.

(iii) Prove that when $\varepsilon = -1$ in (2.7) the linear transformation associated with A is reflection across the line

$$(c-1)x + sy \;=\; 0.$$

When $c = \mathit{cos}(\theta)$ and $s = \mathit{sin}(\theta)$ with $\theta > 0$, this is the line obtained by rotating the x-axis about the origin through an angle of $\theta/2$ counterclockwise. Prove that whenever $\varepsilon = -1$, we have$|A| = 2$ in O_2.

(iv) Let $A_1, A_2 \in O_2$, correspond to reflections about lines ℓ_1, ℓ_2 through the origin. Prove that the product A_2A_1 corresponds to a rotation. Express the angle of rotation for A_2A_1 in terms of the angle (counterclockwise) from ℓ_1 to ℓ_2.

(v) The *special orthogonal group of degree* 2 is

$$SO_2 \;=\; O_2 \cap SL_2(\mathbb{R}) \;=\; \{A \in O_2 \mid det(A) = 1\}, \qquad (2.8)$$

which is a subgroup of O_2. It consists of those A as in (2.7) with $\varepsilon = 1$, i.e., the rotations. Prove that the elements of $O_2 \setminus SO_2$, the reflections (with $\varepsilon = -1$), form a left and right coset SO_2. Deduce that SO_2 is a normal subgroup of O_2 with $|O_2{:}SO_2| = 2$.

Rigid motions of $\mathbb{R}^2$. Let $\mathbb{R}^2 = \{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mid x, y \in \mathbb{R}\}$, the 2-dimensional vector space over $\mathbb{R}$ of column vectors of length 2. We identify $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ with the point (x, y) in the real plane. The usual Euclidean norm on $\mathbb{R}^2$ is given by:

$$\| \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \| \;=\; \sqrt{x^2 + y^2}.$$

So, the distance between $v_1 = \left(\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right)$ is

$$d(v_1, v_2) \;=\; \|v_1 - v_2\|.$$

A *rigid motion* on $\mathbb{R}^2$ is a distance-preserving function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, i.e., $d(f(v_1), f(v_2)) = d(v_1, v_2)$ for all $v_1, v_2 \in \mathbb{R}^2$. Such an f is also called an *isometry* of $\mathbb{R}^2$.

To visualize a rigid motion, imagine a sheet of paper lying flat on a horizontal coordinate plane. The sheet is moved to a new location without distortion. The function from the original location of points on the sheet to their new location is a rigid motion. For example, the motion is a translation if the the sheet is moved laterally without twisting. The motion is a rotation if the sheet is twirled about a point that is held fixed. But we also have a rigid motion if the sheet is turned over. If the sheet is flipped while a line in it is held fixed, the rigid motion is reflection across that line. The next problem shows that all rigid motions of the plane are compositions of translations, rotations, and reflections. See the text by M. Artin [**1**] for a nice discussion of rigid motions.

2.11. *Rigid motions.* Let $\mathcal{RM}$ denote the set of all rigid motions of $\mathbb{R}^2$. Note that if $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a rigid motion, then f sends the vertices of a triangle to the corresponding vertices of a congruent

triangle. Hence, f preserves the angles of a triangle. (This is also evident from the Law of Cosines.)

(i) Prove that $\mathcal{RM}$, with the operation of composition of functions, is a subgroup of $\Sigma(\mathbb{R}^2)$.

(ii) For $w \in \mathbb{R}^2$, the map $\tau_w \colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $\tau_w(v) = v + w$ is called *translation by* w. Clearly, $\tau_w \in \mathcal{RM}$. Let

$$T = \{\tau_w \mid w \in \mathbb{R}^2\},$$

the set of all translations. Prove that T is a subgroup of $\mathcal{RM}$ and that $T \cong \mathbb{R}^2$.

(iii) Let $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a rigid motion. Prove that if v_1 and v_2 are distinct points in $\mathbb{R}^2$, then f maps the line determined by v_1 and v_2 to the line determined by $f(v_1)$ and $f(v_2)$.

(iv) Let $\mathrm{O} = \binom{0}{0}$, the origin in the real plane. Let

$$\mathcal{RM}_0 = \{f \in \mathcal{RM} \mid f(\mathrm{O}) = \mathrm{O}\}.$$

Prove that $\mathcal{RM}_0$ is a subgroup of $\mathcal{RM}$, but not a normal subgroup, and that every element of $\mathcal{RM}_0$ is either a rotation about O or a reflection across a line through O. Deduce that $\mathcal{RM}_0 \cong O_2$, the orthogonal group of degree 2.

(v) Prove that $\mathcal{RM}_0 \cap T = \{1\}$ and that $\mathcal{RM} = T\,\mathcal{RM}_0$. Prove also that T is a normal subgroup of $\mathcal{RM}$.

More generally, for any integer $n > 2$ there is a corresponding group $\mathcal{RM}(\mathbb{R}^n)$ of rigid motions on $\mathbb{R}^n$, and the analogues to (i)–(v) above hold, except that the elements of $\mathcal{RM}(\mathbb{R}^n)_0$ are not always reflections or rotations. These elements are fully decribed in problem 4.105 below.

2.12. *Generalized dihedral groups.* Let G be a group, and let A be a proper subgroup of G. Suppose that every element of $G \setminus A$ has order 2.

(i) Prove that for every $g \in G \setminus A$ and $a \in A$, $gag^{-1} = a^{-1}$.

(ii) Prove that A is an abelian normal subgroup of G.

(iii) Prove that one of (a) or (b) must occur:
   (a) Every element of A has order at most 2. In this case, G is abelian, every nonidentity element of G has order 2, and A could be any subgroup of G.
   (b) A has an element of order exceeding 2. Prove that then G is nonabelian and $|G{:}A| = 2$. In this case G is called a *generalized dihedral group* with distinguished subgroup A.

Example 2.13.

(i) Generalized dihedral groups are generalizations of the usual dihedral groups: For any integer $n \geq 3$, the *dihedral group* D_n is the subgroup of $\mathcal{RM}$ consisting of those rigid motions that send a regular n-sided polygon centered at the origin to itself. Since any map in D_n is determined by its effect on the vertices of the n-gon and must send adjacent vertices to adjacent vertices, D_n has at most $2n$ elements. (There are n choices for where a given vertex can be mapped to, then two choices for where an adjacent vertex is sent; the adjacency condition then determines where the remaining vertices go.) Clearly, D_n contains the rotations through angles of $\frac{2j\pi}{n}$ (radians) for $j = 0, 1, \ldots, n-1$; it also contains the reflections through the n axes of symmetry of the n-gon. Thus, $|D_n| = 2n$, and we have identified all of its elements. The n rotations form a cyclic subgroup of D_n of order n, and every remaining element of D_n is a reflection, so has order 2. Hence, D_n is a generalized dihedral group. For more on D_n, see (2.45) and Example 2.83 below.

(ii) The orthogonal group O_2 is a generalized dihedral group with distinguished subgroup SO_2. (See problem 2.10 above.)

For more on generalized dihedral groups, see problems 2.28, 2.68 (where they are completely classified), and 2.74 below.

Direct products and sums. Let $\{G_i\}_{i \in I}$ be a collection of groups. The (external) *direct product of the* G_i is the Cartesian product $\prod_{i \in I} G_i$ with componentwise multiplication, which is a group. For each $j \in I$, the projection map $\pi_j \colon \prod_{i \in I} G_i \to G_j$ given by sending an element

to its j-component is a surjective group homomorhpism. The direct product is characterized by the following universal mapping property: For any group H and any family of homomorphisms $\alpha_i\colon H \to G_i$ for all $i \in I$, there is a unique homomorphism $\beta\colon H \to \prod_{i\in I} G_i$ such that $\pi_j \circ \beta = \alpha_j$ for each j. The map β is given by $\beta(h) = (\ldots, \underset{j}{\alpha_j(h)}, \ldots)$. For each $j \in I$ there is also an inclusion homomorphism

$$\iota_j\colon G_j \longrightarrow \prod_{i\in I} G_i$$

mapping g_j in G_j to the I-tuple with j-component g_j and i-component 1_{G_i} for $i \neq j$.

If the G_i are all abelian groups, then the *direct sum* of the G_i is

$$\bigoplus_{i\in I} G_i = \Big\{ (\ldots, \underset{i}{g_i}, \ldots) \in \prod_{i\in I} G_i \mid g_i = 1_{G_i} \text{ for all but finitely many } i \in I \Big\}. \tag{2.9}$$

Thus, the direct sum is the subgroup of $\prod_{i\in I} G_i$ generated by the union $\bigcup_{i\in I} \mathit{im}(\iota_i)$. Note that the direct sum is characterized by the following universal mapping property: For any abelian group H and any family of homomorphisms $\gamma_i\colon G_i \to H$, there is a unique homomorphism $\delta\colon \bigoplus_{i\in I} G_i \to H$, such that $\gamma_i = \delta \circ \iota_i$ for all $i \in I$.

2.14. *Internal direct products.* Let G be a group with normal subgroups N_1 and N_2 such that

$$N_1 \cap N_2 = \{1_G\} \quad \text{and} \quad N_1 N_2 = G.$$

Prove that every element of G is expressible uniquely as $n_1 n_2$ with $n_1 \in N_1$ and $n_2 \in N_2$, and that $n_1 n_2 = n_2 n_1$. (Consider $n_1 n_2 n_1^{-1} n_2^{-1}$.) Deduce that the map $G \to N_1 \times N_2$ given by $n_1 n_2 \mapsto (n_1, n_2)$ is a group isomorphism. When this occurs, G is said to be the (*internal*) *direct product* of N_1 and N_2, and we write

$$G = N_1 \times N_2.$$

More generally, for normal subgroups $N_1, \ldots, N_k$ of G, for $k \geq 2$, if $G = N_1 N_2 \ldots N_k$ and $N_j \cap N_1 N_2 \ldots N_{j-1}$ for $j = 2, 3, \ldots, k$ then $G = N_1 \times N_2 \times \ldots \times N_k$.

2.2. Group homomorphisms and factor groups

Recall that subgroup N of a group G is normal in G if $aN = Na$ for every $a \in G$. We sometimes write $N \triangleleft G$ to say that N is normal in G. When this occurs, let

$$G/N = \{aN \mid a \in G\},$$

the set of cosets of N in G. Thus,

$$|G/N| = |G{:}N|.$$

Because of the normality, there is a well-defined group operation on G/N given by

$$(aN) \cdot (bN) = (a \cdot b)N \qquad \text{for all } \ a, b \in G.$$

Then G/N with this operation is called the *factor group* (or quotient group) of G modulo N. The map $\pi\colon G \to G/N$ given by $g \mapsto gN$ is a surjective group homomorphism, called the *canonical projection*.

2.15. Let G be a group, and let $\mathcal{S}$ be a collection of nonempty subsets of G that form a partition of G. Suppose that for each $S, T \in \mathcal{S}$ we have $ST \in \mathcal{S}$, where $ST = \{st \mid s \in S,\ t \in T\}$. Let N be the set in $\mathcal{S}$ that contains 1_G. Prove that N is a normal subgroup of G and that $\mathcal{S}$ consists of the left (= right) cosets of N in G.

2.16. Since $\mathbb{Q}$ is an abelian group with respect to addition, its subgroup $\mathbb{Z}$ is normal in $\mathbb{Q}$. Consider the factor group $\mathbb{Q}/\mathbb{Z}$, whose elements are the cosets $q + \mathbb{Z}$ for $q \in \mathbb{Q}$.

(i) For $r \in \mathbb{Z} \setminus \{0\}$ and $s \in \mathbb{N}$, prove that

$$\left|\tfrac{r}{s} + \mathbb{Z}\right| = s/gcd(r, s).$$

Thus, every element of $\mathbb{Q}/\mathbb{Z}$ has finite order, even though every nonidentity element of $\mathbb{Q}$ has infinite order.

(ii) For any $n \in \mathbb{N}$ prove that $\langle \frac{1}{n} + \mathbb{Z} \rangle$ is the unique cyclic subgroup of $\mathbb{Q}/\mathbb{Z}$ of order n.

(iii) Prove that every subgroup of $\mathbb{Q}/\mathbb{Z}$ generated by finitely many elements is a cyclic group.

Let G and H be groups, and let $\alpha\colon G \to H$ be a group homomorphism. The *image* of α is

$$im(\alpha) = \{\alpha(g) \mid g \in G\}, \tag{2.10}$$

which is a subgroup of H. The *kernel* of α is

$$ker(\alpha) \;=\; \{g \in G \mid \alpha(g) = 1_H\}, \tag{2.11}$$

which is a normal subgroup of G. Note that for any $g \in G$,

$$\alpha^{-1}(\alpha(g)) \;=\; g\,ker(\alpha) \;=\; ker(\alpha)\,g.$$

Hence, α is injective iff $ker(\alpha) = \{1_G\}$.

We now recall the basic theorems about group homomorphisms and isomorphisms. They are essential for the study of groups, though the proofs are easy and can be found in any textbook. There is no general agreement on the numbering of the theorems. Some authors call the Fundamental Homomorphism Theorem the First Isomorphism Theorem.

The *Fundamental Homomorphism Theorem* (FHT) for groups says: Let G, H be groups, and let $\alpha\colon G \to H$ be a group homomorphism. Let N be a normal subgroup of G, and let $\pi\colon G \to G/N$ be the canonical homomorphism. Suppose that $N \subseteq ker(\alpha)$. Then, there is (well-defined) unique induced homomorphism $\beta\colon G/N \to H$ such that $\alpha = \beta \circ \pi$, i.e.,

$$\beta(gN) \;=\; \alpha(g) \qquad \text{for all} \;\; g \in G.$$

Moreover, $im(\beta) = im(\alpha)$ and $ker(\beta) = ker(\alpha)/N$. In particular (taking $N = ker(\alpha)$),

$$G/\,ker(\alpha) \;\cong\; im(\alpha).$$

The *First Isomorphism Theorem* says: Let N be a normal subgroup of group G, and let H be any subgroup of G. Then, HN is a subgroup of G, $H \cap N$ is a normal subgroup of H, and

$$H/(H \cap N) \;\cong\; HN/N.$$

Hence, $|H : H \cap N| = |HN{:}N|$.

The *Second Isomorphism Theorem* says: Let N be a normal subgroup of a group G, and let K be a normal subgroup of G with $K \supseteq N$. Then, there is a well-defined homomorphism $\alpha\colon G/N \to G/K$ given

by $gN \mapsto gK$ for $g \in G$. Also, $ker(\alpha) = K/N$, which is a normal subgroup of G/N, and α induces an isomorphism

$$G/N/K/N \cong G/K.$$

Moreover, every normal subgroup of G/N has the form K/N for some normal subgroup K of G with $K \supseteq N$.

The *Correspondence Theorem* says: Let $\alpha\colon G \to G'$ be a surjective homomorphism of groups. Then there is a one-to-one correspondence between the set $\mathcal{S}$ of subgroups of G containing $ker(\alpha)$ and the set $\mathcal{S}'$ of subgroups of G'. When H in $\mathcal{S}$ corresponds to H' in $\mathcal{S}'$, we have

$$H' = \alpha(H) \quad \text{and} \quad H = \alpha^{-1}(H') = \{h \in G \mid \alpha(h) \in H'\}.$$

Moreover, the correspondence is inclusion-preserving: If H_1, H_2 in $\mathcal{S}$ correspond to H_1', H_2' in $\mathcal{S}'$, then $H_1 \subseteq H_2$ iff $H_1' \subseteq H_2'$. When these inclusions occur, there is also a one-to-one correspondence between the left cosets of H_1 in H_2 and the left cosets of H_1' in H_2'. Likewise for right cosets. Hence, $|H_2{:}H_1| = |H_2'{:}H_1'|$. Furthermore, H_1 is normal in H_2 iff H_1' is normal in H_2', and when this occurs $H_2/H_1 \cong H_2'/H_1'$.

Example 2.17. $\mathbb{Z}_n$. Fix $n \in \mathbb{N}$. The elements $[i]_n = i + n\mathbb{Z}$ of $\mathbb{Z}_n$ for $i \in \mathbb{Z}$ are the cosets of $n\mathbb{Z}$ in $\mathbb{Z}$, and the group operation on $\mathbb{Z}_n$ given by $[i]_n + [j]_n = [i + j]_n$ coincides with the operation on the factor group $\mathbb{Z}/n\mathbb{Z}$ induced by addition on $\mathbb{Z}$. Thus,

$$\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z},$$

and the canonical projection $\pi\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is given by $i \mapsto [i]_n$. Since we know the subgroups of $\mathbb{Z}$ (see Example 2.1) the Correspondence Theorem yields complete information on subgroups of $\mathbb{Z}_n$: Every subgroup of $\mathbb{Z}$ containing $n\mathbb{Z}$ has the form $d\mathbb{Z}$ for $d \in \mathbb{N}$ with $d|n$. For such a d, the Second Isomorphism Theorem yields:

$$\mathbb{Z}_n \,/\, d\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z} = \mathbb{Z}_d.$$

So,

$$|\mathbb{Z}_n : d\mathbb{Z}/n\mathbb{Z}| = |\mathbb{Z}_d| = d;$$

hence, by Lagrange's Theorem,

$$\big|d\mathbb{Z}/n\mathbb{Z}\big| = \big|\mathbb{Z}_n\big|/|\mathbb{Z}_n : d\mathbb{Z}/n\mathbb{Z}| = n/d.$$

The groups $d\mathbb{Z}/n\mathbb{Z}$ are all the subgroups of $\mathbb{Z}_n$, and they are all cyclic, since $d\mathbb{Z}/n\mathbb{Z} = \langle [d]_n \rangle$.

Example 2.18. *Cyclic groups.* Let G be a group, and take any $a \in G$. There is a group homomorphism $\gamma\colon \mathbb{Z} \to G$ given by $\gamma(i) = a^i$ for $i \in \mathbb{Z}$. Clearly, $\mathit{im}(\gamma) = \langle a \rangle$. If $|a| = \infty$, then $\mathit{ker}(\gamma) = \{0\}$, so $\langle a \rangle \cong \mathbb{Z}$. If $|a| = n \in \mathbb{N}$, then $\mathit{ker}(\gamma) = n\mathbb{Z}$, and by the FHT, $\langle a \rangle \cong \mathbb{Z}_n$. Thus, a cyclic group is determined up to isomorphism by its order, and the results about subgroups and factor groups of $\mathbb{Z}$ and $\mathbb{Z}_n$ carry over to all cyclic groups. In particular, every cyclic group of order n is isomorphic to $C_n = \langle e^{2\pi i/n} \rangle$, which is why we call C_n "the" cyclic group of order n. Note that the formula of (1.13),

$$\sum_{d \in \mathbb{N}, d|n} \varphi(d) \;=\; n$$

follows immediately from the subgroup structure of C_n: The cyclic group C_n has order n, each element has order d for some divisor d of n, and there are $\varphi(d)$ elements of order d since they are the generators of the unique cyclic subgroup of C_n of order d.

Example 2.19. *Chinese Remainder Theorem.* For $m, n \in \mathbb{N}$ with $\mathit{gcd}(m, n) = 1$, there is a group homomorphism

$$\theta\colon \mathbb{Z} \to \mathbb{Z}_m \times \mathbb{Z}_n \qquad \text{given by} \qquad j \mapsto ([j]_m, [j]_n)$$

Clearly, $\mathit{ker}(\theta) = m\mathbb{Z} \cap n\mathbb{Z} = mn\mathbb{Z}$. By the FHT, θ induces an isomorphism $\mathbb{Z}_{mn} \to \mathit{im}(\theta)$. Hence,

$$|\,\mathit{im}(\theta)| \;=\; \big|\mathbb{Z}_{mn}\big| \;=\; \big|\mathbb{Z}_m \times \mathbb{Z}_n\big|,$$

so θ is surjective; therefore,

$$\mathbb{Z}_{mn} \;\cong\; \mathbb{Z}_m \times \mathbb{Z}_n.$$

This is a version of the Chinese Remainder Theorem (cf. (1.5)). More generally, the same method (or induction) shows that for $n_1, n_2, \ldots, n_k$ in $\mathbb{N}$ such that $\mathit{gcd}(n_i, n_j) = 1$ whenever $i \neq j$,

$$\mathbb{Z}_{n_1 \ldots n_k} \;\cong\; Z_{n_1} \times \ldots \times Z_{n_k}. \tag{2.12}$$

2.20. Let G be a finite group with $|G| = n$. Suppose that for each divisor d of n there are at most d elements a of G such that $a^d = 1_G$. Prove that G is a cyclic group. (Use formula (1.13).)

2.21. Prove that

$$SO_2 \cong \{z \in \mathbb{C} \mid |z| = 1\} \cong \mathbb{R}/\mathbb{Z},$$

where for $z \in \mathbb{C}$, $|z|$ denotes the absolute value of z. (Note the homomorphism $\mathbb{R} \to \mathbb{C}^*$ given by $r \mapsto e^{2\pi ir}$.)

2.22. Take any $m, n \in \mathbb{N}$. Let $d = gcd(m, n)$ and $\ell = lcm(m, n)$.

(i) Prove that $C_m \times C_n \cong C_\ell \times C_d$.

(ii) Prove that if $C_m \times C_n \cong C_r \times C_s$ with $s \,|\, r$, then $r = \ell$ and $s = d$.

2.23. Let a and b be two elements of finite order in a group G, say $|a| = m$ and $|b| = n$. If $ab = ba$, determine all possible values of $|ab|$. (Clearly, $|ab| \,\big|\, mn$; also, if $gcd(m, n) = 1$, prove that $|ab| = mn$. But, in general, there can be multiple possibilities for $|ab|$, and which ones occur depend on the prime factorization of m and n. For example, if $m = 12$ and $n = 36$, then $|ab| = 9$, 18, or 36, and all these possibilities do occur.)

Note that the assumption that $ab = ba$ is essential here. If a and b do not commute, then $|a|$ and $|b|$ give no information on $|ab|$; see problem 4.112 below.

Automorphisms. Let G be a group. An *automorphism* of G is an isomorphism from G onto G. The *automorphism group of* G is

$$Aut(G) = \{\text{automorphisms of } G\}. \tag{2.13}$$

It is easy to check that $Aut(G)$ is a subgroup of $\Sigma(G)$, with group operation composition of functions. For any $a \in G$, the map $\gamma_a \colon G \to G$ given by $g \mapsto aga^{-1}$, called *conjugation by* a, is an automorphism of G. The maps γ_a for $a \in G$ are called *inner automorphisms* of G. The map $\beta \colon G \to Aut(G)$ given by $a \mapsto \gamma_a$ is a group homomorphism. Its image is the group inner automorphisms of G,

$$\mathcal{I}nn(G) = \{\gamma_a \mid a \in G\}, \tag{2.14}$$

a subgroup of $Aut(G)$. In fact, $\mathcal{I}nn(G)$ is a normal subgroup of $Aut(G)$. Note also that $ker(\beta)$ is the center $Z(G)$ of G (see (2.6)), which is a normal and abelian subgroup of G. The Fundamental Homomorphism Theorem shows that β induces an isomorphism

$$G/Z(G) \cong \mathcal{I}nn(G).$$

2.24. Fix $n \in \mathbb{N}$. We determine the automorphism group of the cyclic group C_n of order n.

(i) For each $j \in \mathbb{Z}$ there is a group homomorphism $\mu_j : C_n \to C_n$ given by $b \mapsto b^j$ for all $b \in \mathbb{C}_n$. Prove that the μ_j for $j \in \mathbb{Z}$ are all the group homomorphisms of C_n to itself, and that $\mu_j = \mu_\ell$ iff $j \equiv \ell \ (mod\ n)$.

(ii) Recall the group $\mathbb{Z}_n^*$ described in (2.3) and (2.4). Deduce from part (i) that

$$Aut(C_n) \cong \mathbb{Z}_n^*. \tag{2.15}$$

2.25. Let G be a finite group and let N be a normal subgroup of G. Let $n = |N|$ and $m = |G{:}N|$, and assume that $gcd(m, n) = 1$. Let H be any subgroup of G. Since $|H| \,\big|\, |G| = mn$ with m and n relatively prime, there are unique positive integers k, ℓ such that $|H| = k\ell$ with $k|n$ and $\ell|m$. Prove that $|H \cap N| = k$ and $|H : H \cap N| = \ell$.

2.26. Let H and K be finite groups with $gcd\big(|H|, |K|\big) = 1$.

(i) Prove that every subgroup of $H \times K$ has the form $A \times B$ where A is a subgroup of H and B is a subgroup of K. (Hint: Use the preceding problem.)

(ii) Prove that $Aut(H \times K) \cong Aut(H) \times Aut(K)$.

Note that the assumption that $gcd\big(|H|, |K|\big) = 1$ is essential for this problem. For example, if p is prime, then $C_p \times C_p$ has $p + 1$ subgroups of order p, since it has $p^2 - 1$ elements of order p and each subgroup of order p has a different subset of $p - 1$ of those elements. But only two of those subgroups has the form $A \times B$ for A, B subgroups of C_p. Furthermore, $Aut(C_p) \times Aut(C_p) \cong \mathbb{Z}_p^* \times \mathbb{Z}_p^*$, which has order $(p-1)^2$, while $Aut(C_p \times C_p) \cong GL_2(\mathbb{Z}_p)$, which has order $(p^2 - 1)(p^2 - p)$ (see (2.61) below).

The preceding problem applies to cyclic groups, yielding (with (2.15) and the Chinese Remainder Theorem as in Example 2.19): For $m, n \in \mathbb{N}$ with $gcd(m, n) = 1$,

$$\begin{aligned} \mathbb{Z}_{mn}^* &\cong Aut(C_{mn}) \cong Aut(C_m \times C_n) \\ &\cong Aut(C_m) \times Aut(C_n) \cong \mathbb{Z}_m^* \times \mathbb{Z}_n^*. \end{aligned} \tag{2.16}$$

This provides a way to verify the product formula (1.12) for Euler's φ-function: Whenever $gcd(m,n) = 1$,

$$\varphi(mn) \;=\; \left|\mathbb{Z}^*_{mn}\right| \;=\; \left|\mathbb{Z}^*_m \times \mathbb{Z}^*_n\right| \;=\; \varphi(m)\,\varphi(n).$$

2.27. It is known (see §2.10 below) that every finite abelian group is a direct product of cyclic groups. This problem gives such a direct product decomposition for $\mathbb{Z}^*_n$ for $n \in \mathbb{N}$ with $n \geq 2$. For distinct prime numbers $p_1, \ldots, p_k$ and positive integers $r_1, \ldots, r_k$, (2.16) yields

$$\mathbb{Z}^*_{p_1^{r_1}\ldots p_k^{r_k}} \;\cong\; \mathbb{Z}^*_{p_1^{r_1}} \times \ldots \times \mathbb{Z}^*_{p_k^{r_k}}.$$

Thus, it suffices to consider $\mathbb{Z}^*_{p^r}$ for p prime.

(i) Let p be an odd prime number. Because $\mathbb{Z}^*_p = \mathbb{Z}_p \setminus \{[0]_p\}$, is the multiplicative group of the finite field $\mathbb{Z}_p$, it is known that $\mathbb{Z}^*_p$ is a cyclic group. (See problem 3.32 below.) So $\mathbb{Z}^*_p \cong C_{p-1}$. Building on this, prove that for $r \geq 2$,

$$\mathbb{Z}^*_{p^r} \;\cong\; C_{p-1} \times C_{p^{r-1}}.$$

(Hint: To find an element of order p^{r-1} in $\mathbb{Z}^*_{p^r}$, prove and use the identity

$$(1+p)^{p^k} \;\equiv\; 1 + p^{k+1} \;(mod\; p^{k+2}),$$

for all $k \in \mathbb{N}$.)

(ii) Prove that for $r \geq 3$,

$$\mathbb{Z}^*_{2^r} \;\cong\; C_{2^{r-2}} \times C_2.$$

(Hint: Prove and use the identity

$$5^{2^k} \;\equiv\; 1 + 2^{k+2} \;(mod\; 2^{k+3}),$$

for all $k \in \mathbb{N}$.)

2.28. *Subgroups of generalized dihedral groups.* Let G be a generalized dihedral group with distinguished subgroup A, as in problem 2.12.

(i) Let B be a subgroup of A. Prove that B is a normal subgroup of G. Prove that every element of $G/B \setminus A/B$ has order 2. Hence, either G/B is abelian with every element of order at most 2 or G/B is generalized dihedral with distinguished subgroup A/B.

(ii) With B as in part (i), take any $g \in G \setminus A$. Set

$$D = B \cup gB.$$

Prove that D is a subgroup of G with $D \cap A = B$. Note that since every element of $D \setminus B$ has order 2 (as $D \setminus B \subseteq G \setminus A$), either D is abelian with every element of order at most 2 or D is generalized dihedral with distinguished subgroup B. Prove further that D is normal in G iff G/B is abelian (which occurs iff every element of A/B has order at most 2).

(iii) Prove that the subgroups B as in part (i) and D as in part (ii) are all the subgroups of G.

(iv) Prove that if C is a subgroup of G such that every element of $G \setminus C$ has order 2, then $C = A$. Thus, A is the only distinguished subgroup of the generalized dihedral group G.

2.29. Let G be a group, and let H be a subgroup of G. Suppose that $|G{:}H| = n < \infty$. Let $a \in G$. Prove that there is a $j \in \mathbb{N}$ with $j \leq n$ such that $a^j \in H$. Give an example to show that the least such j need not divide n. (Note, however, that if H is normal in G, then the least j is the order of aH in the group G/H, which divides $|G/H| = n$.)

2.30. Let G be a group with only finitely many different subgroups. Prove that G has finite order.

2.3. Group actions

Let G be a group and let S be a set. A (*left*) *group action* of G on S is a pairing $G \times S \to S$, with the image of (g, s) in S denoted $g \cdot s$, such that for all $g, h \in G$ and $s \in S$,

$$g \cdot (h \cdot s) = (gh) \cdot s \qquad \text{and} \qquad 1_G \cdot s = s.$$

For $s \in S$, the *orbit of* s is $\mathcal{O}(s) = \{g \cdot s \mid g \in G\}$. The *stabilizer of* s (or isotropy group of s) is

$$G_s = \{g \in G \mid g \cdot s = s\},$$

a subgroup of G. It is easy to check that the map from the left cosets of G_s in G to $\mathcal{O}(s)$ given by $gG_s \mapsto g \cdot s$ is a well-defined one-to-one

correspondence; this yields the *orbit equation*:

$$|G{:}G_s| \;=\; \big|\mathcal{O}(s)\big|. \tag{2.17}$$

The orbits for the action form a partition of S, since they are the equivalence classes for the equivalence relation $\sim$ on S defined by

$$s \sim s' \;\text{ just when there is } g \in G \text{ with } s' = g \cdot s.$$

Hence, distinct orbits are disjoint, and if $\{\mathcal{O}_i\}_{i\in I}$ is the set of all the different orbits, then

$$|S| \;=\; \sum_{i\in I} |\mathcal{O}_i|. \tag{2.18}$$

The *fixed-point subset of* S for the group action is

$$\mathcal{F}(S) \;=\; \{s \in S \mid \mathcal{O}(s) = \{s\}\} \;=\; \{s \in S \mid G_s = G\}. \tag{2.19}$$

The *kernel of the action* is

$$\mathcal{K}_S(G) \;=\{g \in G \mid g \cdot s = s \text{ for all } s \in S\} \;=\; \bigcap_{s\in S} G_s, \tag{2.20}$$

which is a normal subgroup of G (see the following problem).

2.31. Let G be a group and S a set.

(i) Suppose that there is a group action of G on S. For any $g \in G$ let $\tau(g)\colon S \to S$ be the function given by $s \mapsto g \cdot s$, for all $s \in S$. Prove that $\tau(g)$ is bijective; hence, $\tau(g) \in \Sigma(S)$. Prove that the resulting map $\tau\colon G \to \Sigma(S)$ is a group homomorphism. Note also that $\mathcal{K}_S(G) = ker(\tau)$.

(ii) Conversely, let $\beta\colon G \to \Sigma(S)$ be any group homomorphism. Define a pairing $G \times S \to S$ by $(g, s) \mapsto \beta(g)(s)$. Prove that this pairing gives a group action of G on S. Thus, group actions of G on S are equivalent to homomorphisms of G to $\Sigma(S)$.

Example 2.32. *Group action on cosets.* Let H be a subgroup of a group G, and let S be the set of left cosets,

$$S \;=\; \{aH \mid a \in G\}.$$

There is a well-defined group action of G on S given by

$$g \cdot (aH) \;=\; gaH \qquad \text{for all } \; g, a \in G.$$

This is called the *left action* (or *left translation action*) of G on S. For any $a \in G$, we have $\mathcal{O}(aH) = S$, $G_{aH} = aHa^{-1}$, and

$$\mathcal{K}_S(G) \;=\; \bigcap_{a\in G} aHa^{-1},$$

which is called the *core* of H. Note that $\mathcal{K}_S(G)$ is the largest subgroup of H which is a normal subgroup of G. If $|G{:}H| = n < \infty$, then problem 2.31(i) shows that there is a homomorphism $\tau\colon G \to \Sigma(S)$ with kernel $\mathcal{K}_S(G)$. Hence,

$$|G{:}\mathcal{K}_S(G)| \;=\; |\,\mathit{im}(\tau))|\,\big|\,n!. \tag{2.21}$$

If K is another subgroup of G, then the action of G on S restricts to an action of K on S. For this action, the orbit of $H \in S$ is the set of left cosets of H in KH, and the stabilizer of H is $K \cap H$. The orbit equation yields $|KH|/|H| = |K : K \cap H|$. Thus, when $|K \cap H| < \infty$,

$$|KH| \;=\; |K|\,|H|/|K \cap H|. \tag{2.22}$$

If we take $H = \{1_G\}$, then $S = G$, and the left action of G on G induces a homomorphism $\tau\colon G \to \Sigma(G)$ by problem 2.31(i), and τ is clearly injective. This yields *Cayley's Theorem*: Every group is isomorphic to a subgroup of a symmetric group.

2.33. Suppose group G acts on two sets S and T. We say that the two group actions are *equivalent* if there is a bijective map $f\colon S \to T$ such that $g \cdot (f(s)) = f(g \cdot s)$ for all $g \in G$ and $s \in S$. Suppose that the action of G on S is *transitive*, i.e., there is only one orbit, which is all of S. Prove that then for any $s \in S$ the action of G on S is equivalent to the left action of G on the cosets of G_s.

Example 2.34. *Conjugation action.* Let G be a group. There is a group action of G on G by conjugation: For $g, a \in G$, set

$$g \cdot a \;=\; gag^{-1}.$$

For $a \in G$, the orbit of a under this action is the *conjugacy class of* a,

$$\mathcal{C}\ell(a) \;=\; \{gag^{-1} \mid g \in G\}. \tag{2.23}$$

The elements of $\mathcal{C}\ell(a)$ are called *conjugates* of a. The stabilizer of a is the *centralizer of* a *in* G,

$$C_G(a) \;=\{g \in G \mid ga = ag\}, \tag{2.24}$$

which is a subgroup of G. The orbit equation (2.17) yields

$$\big|\mathcal{C}\ell(a)\big| = |G{:}C_G(a)|. \tag{2.25}$$

The kernel of the action is the center $Z(G)$ of G (see (2.6)). Note that $Z(G)$ is also the fixed set for this action. If G is a finite group, let $\mathcal{C}\ell(a_1), \mathcal{C}\ell(a_2), \ldots \mathcal{C}\ell(a_k)$ be the distinct conjugacy classes of G containing more than one element. Since the the conjugacy classes form a partition of G, equations (2.18) and (2.25) yield the *Class Equation*:

$$|G| = |Z(G)| + \sum_{i=1}^{k} |G{:}C_G(a_i)|. \tag{2.26}$$

G also acts by by conjugation on the collection $\mathcal{S}$ of subsets of G: If $T \in \mathcal{S}$ and $g \in G$, then set

$$g \cdot T = gTg^{-1}.$$

For a subgroup H of G, the orbit of H is the set of *conjugates of* H, i.e., $\{gHg^{-1} \mid g \in G\}$. Note that since conjugation by g is an automorphism of G every conjugate of H is a subgroup of G isomorphic to H. The stabilizer of H for the conjugacy action is the *normalizer of* H *in* G,

$$N_G(H) = \{g \in G \mid gHg^{-1} = H\} = \{g \in G \mid gH = Hg\}, \tag{2.27}$$

which is the largest subgroup of G containing H as a normal subgroup. The orbit equation (2.17) shows that

$$|G{:}N_G(H)| = \text{ the number of conjugates of } H \text{ in } G.$$

2.35.

(i) Let G be a finite group, and let H be a proper subgroup of G. Prove that $G \neq \bigcup_{g \in G} gHg^{-1}$.

(ii) Prove that the conclusion of part (i) still holds if G is infinite and $|G{:}H| < \infty$.

Note that the requirement in part (ii) that $|G{:}H| < \infty$ cannot be dropped. For example, let $G = GL_n(\mathbb{C})$ for $n \geq 2$, and let B be the subgroup of upper triangular matrices in G. Then, since $\mathbb{C}$ is algebraically closed (see p. 195 and problem 5.101), the triangulability theorem in linear algebra (see (4.60)) says that $G = \bigcup_{g \in G} gBg^{-1}$.

2.36. Let G be a finite group, and let p be the least prime number dividing $|G|$. If H is a subgroup of G with $|G{:}H| = p$ prove that H is a normal subgroup of G. (This generalizes the elementary fact that if $|G{:}H| = 2$ then $H \triangleleft G$.)

2.37. Let group G act on a set S, and let N be a normal subgroup of G. The action of G restricts to an action of N on S. Let $\{\mathcal{O}_N(s_i)\}_{i\in I}$ be the collection of distinct orbits for the action of N on S.

(i) Prove that there is a well-defined action of G/N on the set $\{\mathcal{O}_N(s_i)\}_{i\in I}$ of N-orbits of S given by

$$gN \cdot \mathcal{O}_N(s_i) \;=\; \mathcal{O}_N(g \cdot s_i).$$

(ii) Deduce that for $s, t \in S$, if s and t lie in the same G-orbit of S, then $|\mathcal{O}_N(s)| = |\mathcal{O}_N(t)|$.

2.4. Symmetric and alternating groups

Recall that for $n \in \mathbb{N}$, the n-th symmetric group S_n is the group of bijective functions from $\{1, 2, \ldots, n\}$ to itself, with group operation composition of functions. Thus,

$$|S_n| \;=\; n!.$$

The elements of S_n are called *permutations* of the numbers $1, 2, \ldots, n$. We write id_n for the identity map on $\{1, 2, \ldots, n\}$, which is the identity element of S_n. For $\sigma \in S_n$, the *support of* σ is

$$supp(\sigma) \;= \{j \in \{1, 2, \ldots, n\} \mid \sigma(j) \neq j\}. \tag{2.28}$$

Observe that $|\,supp(\sigma)| \neq 1$. We say that $\sigma, \tau \in S_n$ are *disjoint* if $supp(\sigma) \cap supp(\tau) = \varnothing$. Note that if σ and τ are disjoint, then $\sigma\tau = \tau\sigma$, $supp(\sigma\tau) = supp(\sigma) \cup supp(\tau)$, and since σ^i and τ^j are disjoint for all $i, j \in \mathbb{N}$,

$$|\sigma\tau| \;=\; lcm(|\sigma|, |\tau|).$$

Cycles. For $2 \leq k \leq n$, take any distinct numbers $i_1, i_2, \ldots, i_k$ in $\{1, 2, \ldots, n\}$. Let $(i_1 \;\; i_2 \;\; \ldots \;\; i_k)$ denote the function on $\{1, 2, \ldots, n\}$

defined by

$$i_1 \mapsto i_2,\ i_2 \mapsto i_3,\ \ldots,\ i_j \mapsto i_{j+1},\ \ldots,\ i_{k-1} \mapsto i_k,\ i_k \mapsto i_1,$$
$$\ell \mapsto \ell \quad \text{if} \quad \ell \notin \{i_1, i_2, \ldots, i_k\}.$$

Clearly, this map lies in S_n. Such a map is called a *k-cycle* (or a *cycle of length k*). A 2-cycle is also called a *transposition*. Note that

$$(i_1\ i_2\ \ldots\ i_k) = (i_2\ i_3\ \ldots\ i_k\ i_1) = \ldots$$
$$= (i_j\ i_{j+1}\ \ldots\ i_k\ i_1\ \ldots\ i_{j-1}) = \ldots = (i_k\ i_1\ \ldots\ i_{k-1}).$$

Note that if ψ is a k-cycle, then $|\psi| = k$. Also, every k-cycle is a product of $k-1$ transpositions:

$$(i_1\ i_2\ \ldots\ i_k) = (i_1\ i_2)\,(i_2\ i_3)\ldots(i_j\ i_{j+1})\ldots(i_{k-1}\ i_k). \quad (2.29)$$

The *cycle decomposition theorem* says that every nonidentity permutation σ in S_n is expressible as a product of one or more pairwise disjoint cycles, and the cycles appearing in the product are uniquely determined. The cycles are all obtainable as follows: Take any $j \in supp(\sigma)$ and let k be the least positive integer such that $\sigma^k(j) = j$. One shows that such a k exists and that $j, \sigma(j), \sigma^2(j), \ldots, \sigma^{k-1}(j)$ are all distinct, and that the k-cycle

$$(j\ \sigma(j)\ \sigma^2(j)\ \ldots\ \sigma^{k-1}(j))$$

appears in the cycle decomposition of σ. The next problem gives an approach to proving the cycle decomposition theorem via a group action.

2.38. *Disjoint cycle decomposition.* Let $\sigma \in S_n$ for $n \geq 2$, with $\sigma \neq id_n$. There is a group action of $\langle\sigma\rangle$ on $\{1, 2, \ldots, n\}$ given by $\rho \cdot j = \rho(j)$ for $j = 1, 2, \ldots, n$. (This is the group action associated with the inclusion homomorphism $\iota\colon \langle\sigma\rangle \to S_n$ as in problem 2.31.) Let $\mathcal{O}_1, \ldots, \mathcal{O}_s$ be those orbits of $\{1, 2, \ldots, n\}$ with $|\mathcal{O}_j| > 1$ for this action. So, $supp(\sigma) = \bigcup_{j=1}^{s} \mathcal{O}_i$, a disjoint union.

(i) For $j = 1, 2, \ldots, s$, let $\psi_j \in S_n$ be defined by

$$\psi_j(i) = \begin{cases} \sigma(i), & \text{if } i \in \mathcal{O}_j, \\ i & \text{if } i \notin \mathcal{O}_j. \end{cases}$$

Prove that ψ_j is an $|\mathcal{O}_j|$-cycle in S_n with $supp\,\psi_j = \mathcal{O}_j$. Thus, ψ_i and ψ_j are disjoint whenever $i \neq j$.

(ii) Prove that $\sigma = \psi_1\psi_2 \dots \psi_s$. This formula for σ is the *disjoint cycle decomposition* of σ.

(iii) Prove the uniqueness of the disjoint cycle decomposition of σ, i.e., prove that if $\sigma = \rho_1 \dots \rho_\ell$ with the ρ_i pairwise disjoint cycles, then $\ell = s$ and $\{\rho_1, \dots, \rho_\ell\} = \{\psi_1, \dots, \psi_s\}$.

2.39. Let $\psi = (i_1\ i_2\ \dots\ i_k)$ be any k-cycle in S_n.

(i) Prove that for any $\beta \in S_n$,

$$\beta\psi\beta^{-1} = \big(\beta(i_1)\ \beta(i_2)\ \dots\ \beta(i_k)\big).$$

It follows that the conjugacy class of ψ in S_n is the set of all k-cycles in S_n. It then follows that for any σ in S_n, the conjugacy class of σ is the set of all τ in S_n such that τ has the same number of k-cycles in its disjoint cycle decomposition as σ has, for each $k \in \mathbb{N}$.

(ii) For the centralizer, prove that

$$C_{S_n}(\psi) = \{\psi^i\gamma \mid i \in \mathbb{N} \text{ and } \gamma \in S_n \text{ is disjoint from } \psi\}.$$

(iii) Prove that $|N_{S_n}(\langle\psi\rangle){:}C_{S_n}(\psi)| = \varphi(n)$.

2.40. Take any integer $n \geq 2$.

(i) Prove that $\langle(1\ 2), (1\ 3), \dots, (1\ j), \dots (1\ n)\rangle = S_n$.

(ii) Prove that $\langle(1\ 2), (2\ 3), \dots, (j\ \ j+1), \dots (n-1\ \ n)\rangle = S_n$.

(iii) Let $\tau_1, \tau_2, \dots, \tau_k$ be transpositions in S_n. prove that

$$\text{if } \langle\tau_1, \tau_2, \dots, \tau_k\rangle = S_n, \text{ then } k \geq n-1.$$

2.41. Let ψ be the n-cycle $(1\ 2\ \dots\ i\ \ i+1\ \dots\ n)$, in S_n for $n \geq 2$.

(i) Prove that $\langle\psi, (1\ 2)\rangle = S_n$.

(ii) Determine those j with $2 \leq j \leq n$ for which $\langle\psi, (1\ j)\rangle = S_n$.

(iii) For each i with $2 \leq i \leq n$, determine $\big|\langle\psi, (1\ i)\rangle\big|$.

2.42. Prove that for $n \geq 3$ and $n \neq 6$,

$$Aut(S_n) = \mathcal{I}nn(S_n) \cong S_n.$$

(Hint: First determine the number of products of k disjoint transpositions in S_n for $k \in \mathbb{N}$.)

The case $n = 6$ is a genuine exception here: It is known that $\big|Aut(S_6)\big| = 2\big|S_6\big|$. For more about $Aut(S_6)$, see the paper by Lam and Leep [**14**].

Sign of a permutation. For any $n \in \mathbb{N}$ with $n \geq 2$, let $\sigma \in S_n$. The *sign of* σ is defined to be

$$sgn(\sigma) \;=\; \prod_{1\leq i<j\leq n} \tfrac{\sigma(j)-\sigma(i)}{j-i} \;\in \{1,-1\}. \tag{2.30}$$

The product is take over all pairs of integers i, j with $1 \leq i < j \leq n$. Then, σ is called an *even permutation* if $sgn(\sigma) = 1$, and an *odd permutation* if $sgn(\sigma) = -1$. Since $\frac{\sigma(j)-\sigma(i)}{j-i} = \frac{\sigma(i)-\sigma(j)}{i-j}$, another description of the sign is

$$sgn(\sigma) \;=\; \prod_{\{i,j\}\in\mathcal{S}} \tfrac{\sigma(j)-\sigma(i)}{j-i},$$

where $\mathcal{S}$ is the set of all 2-element subsets of $\{1, 2, \ldots, n\}$. From this formula it is easy to check that

$$sgn(\sigma\rho) \;=\; sgn(\sigma)\, sgn(\rho) \qquad \text{for all } \;\sigma, \rho \in S_n,$$

so the function $sgn\colon S_n \to \{1, -1\}$ is a group homomorphism. For the transposition $(1\ \ 2)$ a short calculation shows that $sgn\big((1\ \ 2)\big) = -1$. Since any transposition τ is a conjugate of $(1\ \ 2)$ and $\{\pm 1\}$ is abelian, we have $sgn(\tau) = -1$. Furthermore, as any k-cycle ψ in S_n is a product of $k-1$ transpositions (see (2.29)), $sgn(\psi) = (-1)^{k-1}$. Thus, for any σ in S_n,

$$sgn(\sigma) \;=\; (-1)^m,$$

where m is the number of cycles of even length in the disjoint cycle decomposition of σ. Since S_n is generated by transpositions, σ is an even (resp. odd) permutation iff σ is expressible as a product of an even (resp. odd) number of transpositions.

The *n-th alternating group* is

$$A_n \;=\; ker(sgn) \;=\; \{\,\text{even permutations in } S_n\,\}. \tag{2.31}$$

Thus, A_n is a normal subgroup of S_n with $|S_n\!:\!A_n| = |im(sgn)| = 2$; so,

$$\big|A_n\big| = n!/2.$$

Since every product of two transpositions is expressible as a product of 3-cycles, A_n is generated by 3-cycles. It is known that for $n \geq 5$ or $n = 3$ the group A_n is *simple*, i.e., it is nontrivial but has no nontrivial proper normal subgroup. See any algebra text for a proof. (A_4 is not simple since the *Klein* 4*-group*

$$\mathcal{K}_4 = \{1_{A_4}, (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4)(2\ 3)\} \tag{2.32}$$

is normal in A_4.)

2.43. *Uniqueness of sgn.*

(i) For $n \geq 2$, let $\gamma\colon S_n \to \{1, -1\}$ be a surjective group homomorphism. Prove that $\gamma = sgn$.

(ii) Deduce that A_n is the only subgroup of S_n of index 2.

2.44. Prove that for any $n \in \mathbb{N}$, the symmetric group S_n is isomorphic to a subgroup of A_{n+2}. (It then follows from Cayley's Theorem that every finite group is isomorphic to a subgroup of some alternating group.)

Note, however, that S_n is not isomorphic to a subgroup of A_{n+1}, for $n \geq 1$. See problem 2.100(iv) below.

2.45. Prove that for $n \geq 5$ or $n = 3$, the only normal subgroups of S_n are S_n, A_n, and $\{1_{S_n}\}$.

2.46. Let H be a subgroup of S_n with $n \geq 5$.

(i) Suppose that $1 < |S_n{:}H| < n$. Prove that $H = A_n$.

(ii) Suppose $|S_n{:}H| = n$. Prove that $H \cong S_{n-1}$. (If $n \neq 6$, one can show further, using that $Aut(S_n) \cong S_n$, that $H = \Sigma(S)$ for some subset S of $\{1, 2, \ldots, n\}$ with $|S| = n - 1$. Here, we are identifying $\Sigma(S)$ with the subgroup of S_n of permutations mapping each element of $\{1, 2, \ldots, n\} \setminus S$ to itself. The case $n = 6$ is exceptional: in fact, S_6 has 12 subgroups of index 6.)

2.5. p-groups

Let p be a prime number. A *p-group* is a finite group whose order is a power of p.

2.47. Let G be a p-group acting on a finite set S. Let $\mathcal{F}(S)$ be the fixed-point subset of S as in (2.19). Prove that

$$\big|\mathcal{F}(S)\big| \equiv |S| \ (mod\ p).$$

2.48. *Cauchy's Theorem.* Let G be a finite group and let p be a prime with $p \,\big|\, |G|$. *Cauchy's Theorem* says that G then contains an element of order p, (hence, G has a subgroup of order p). This problem gives a proof of Cauchy's Theorem. (It is also provable using the Class Equation.) Let

$$S \,=\, \{(a_1, a_2, \ldots, a_p) \mid \text{ each } a_i \in G \text{ and } a_1 a_2 \ldots a_p = 1_G\}.$$

Let $\mathbb{Z}_p$ act on S by cyclic permutations, i.e., for $0 \le i \le p-1$,

$$[i]_p \cdot (a_1, a_2, \ldots, a_p) \,=\, (a_{i+1}, a_{i+2}, \ldots, a_{p-1}, a_p, a_1, a_2, \ldots, a_i).$$

(i) Prove that this is a well-defined group action and that

$$|S| \,=\, |G|^{p-1}.$$

(ii) Use this group action and the preceding problem to prove Cauchy's Theorem.

2.49. Let G be a p-group. Use problem 2.47 to prove that $|Z(G)| \ge p$. (This is also provable using the Class Equation (2.26).)

Maximal subgroups. Let G be a group. A *maximal subgroup* of G is a maximal proper subgroup. That is, a subgroup M of G is maximal when $M \subsetneqq G$ and there is no subgroup H of G with $M \subsetneqq H \subsetneqq G$. Of course, nontrivial finite groups have maximal subgroups, but an infinite group need not have any maximal subgroups, as the next problem illustrates.

2.50.

(i) Let A be an abelian group. Prove that a subgroup B of A is maximal iff $|A{:}B| = p$ for some prime number p.

(ii) Prove that $\mathbb{Q}$ has no maximal subgroup.

2.51. *Frattini subgroup.* Let G be a nontrivial finite group. The *Frattini subgroup of* G, denoted $\mathcal{D}(G)$ is the intersection of all the maximal subgroups of G. Since conjugates of maximal subgroups are maximal, $\mathcal{D}(G)$ is a normal subgroup of G. For any $a_1, \ldots, a_n$ in G, prove that

$$\langle a_1, a_2, \ldots, a_n \rangle = G \text{ iff } \langle a_1\mathcal{D}(G), a_2\mathcal{D}(G), \ldots, a_n\mathcal{D}(G) \rangle = G/\mathcal{D}(G).$$

2.52. Let P be a p-group, and let H be a proper subgroup of P.

(i) Prove that $N_P(H) \supsetneqq H$.

(ii) Deduce that if H is a maximal subgroup of G then H is normal in P and $|P{:}H| = p$, so $P/H \cong C_p$.

2.53. *Elementary abelian p-groups.* For any prime p, let P be a finite abelian group in which every nonidentity element has order p. Such a P is called an *elementary abelian p-group.* Cauchy's Theorem shows that P is a actually a p-group. Suppose that $|P| = p^n$.

(i) Prove that $P \cong \prod_{i=1}^{n} C_p$.

(ii) Prove that P can be generated by n suitably chosen elements, but not by any subset with fewer than n elements. Prove further that any generating set of P has a subset with n elements that already generates P.

(In fact, any elementary abelian p-group P can be viewed as a vector space over the field $\mathbb{Z}_p$, and the subgroups coincide with the $\mathbb{Z}_p$-subspaces. See problem 4.3 below. If $|P| = p^n$, then P is an n-dimensional $\mathbb{Z}_p$-vector space.)

2.54. Let P be an elementary abelian p-group. Prove that

$$\mathcal{D}(P) = \{1_P\}.$$

2.55. Let P be a p-group.

(i) Prove that $P/\mathcal{D}(P)$ is an elementary abelian p-group. (It follows by problems 2.51 and 2.53 that every generating set of P has at least $log_p|P/\mathcal{D}(P)|$ elements; Moreover, every generating set of P has a subset with $log_p|P/\mathcal{D}(P)|$ elements that already generates P.)

(ii) Prove that

$$\mathcal{D}(P) \;=\; \langle \{a^p, aba^{-1}b^{-1} \mid a, b \in G\} \rangle.$$

2.56. Let G be a group and let H be a subgroup of $Z(G)$, so H is a normal subgroup of G. Prove that if G/H is cyclic then G is abelian.

2.57. Let p be prime number. Let G be group with $|G| = p^2$.

(i) Prove that G is abelian. (Hint: Apply the preceding problem.)

(ii) Prove that $G \cong C_{p^2}$ or $G \cong C_p \times C_p$.

2.58. *Heisenberg group.* Let R be any ring, and let $1 = 1_R$. The *Heisenberg group of* R is defined to be

$$\mathcal{H}(R) \;=\; \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \;\middle|\; a, b, c \in R \right\} \;\subseteq\; M_3(R). \tag{2.33}$$

Note that every element of $\mathcal{H}(R)$ has the form $I + N$ where $I = I_3$ is the identity matrix in $M_3(R)$ and N is upper triangular with 0's on the main diagonal. Since $N^3 = 0$, we have

$$I \;=\; I^3 + N^3 \;=\; (I+N)(I-N+N^2) \;=\; (I-N+N^2)(I+N),$$

which shows that $I+N$ has a multiplicative inverse in $\mathcal{H}(R)$. It follows easily that $\mathcal{H}(R)$ is a subgroup of $GL_3(R)$. Now, let p be a prime number with $p \geq 3$; consider the case where $R = \mathbb{Z}_p$ (which is a ring, see p. 74 below). Prove that $\mathcal{H}(\mathbb{Z}_p)$ is a nonabelian group of order p^3 in which every nonidentity element has order p. (Recall (1.16).) Prove also that $\mathcal{H}(\mathbb{Z}_2) \cong D_4$.

2.6. Sylow subgroups

Sylow Theorems. Let G be a finite group and let p be a prime number dividing $|G|$. Write $|G| = p^a b$ with $a, b \in \mathbb{N}$ and $p \nmid b$. A *p-Sylow subgroup* of G is a subgroup of order p^a. We now recall the *Sylow Theorems* on Sylow subgroups, which are key results for the study of finite groups. See almost any algebra text for proofs. Let G be a group of order $p^a b$ as above. Then:

(i) G has a p-Sylow subgroup.

(ii) If P is a p-Sylow subgroup of G and Q is any subgroup of G that is a p-group, then $Q \subseteq gPg^{-1}$ for some $g \in G$. Hence, the p-Sylow subgroups of G are the conjugates of P in G.

(iii) Let n_p be the number of p-Sylow subgroups of G. Then,

$$n_p \equiv 1 \ (mod\ p).$$

(Also, $n_p | b$, since by (ii), $n_p = |G{:}N_G(P)|$, which divides $|G{:}P| = b$, as $N_G(P) \supseteq P$.)

2.59. Let G be a group of order pqr, where p, q, and r are prime numbers with $p < q < r$. Prove that G has a normal (hence unique) r-Sylow subgroup.

2.60. Let G be a group of order p^2q^2 for distinct primes p and q.

(i) If $|G| \neq 36$, prove that G has a normal Sylow subgroup.

(ii) If $|G| = 36$, prove that G has a normal Sylow subgroup.

2.61. Let G be a finite group and let P be a p-Sylow subgroup of G. Let K be a normal subgroup of G.

(i) Prove that if $p \,\big|\, |K|$, then $P \cap K$ is a p-Sylow subgroup of K, and that all p-Sylow subgroups of K arise this way.

(ii) Prove that if $p \,\big|\, |G{:}K|$, then PK/K is a p-Sylow subgroup of G/K, and that all p-Sylow subgroups of G/K arise this way.

2.62. Let P be a p-Sylow subgroup of a group G, and let H be a subgroup of G with $H \supseteq N_G(P)$. Prove that $N_G(H) = H$.

2.63. Let G be a finite group with a normal subgroup K, and let P be a p-Sylow subgroup of K. Prove that $G = KN_G(P)$.

2.7. Semidirect products of groups

2.64. *Internal semidirect product.* Let G be a group and let N and H be subgroups of G with N normal in G. Suppose that

$$NH = G \quad \text{and} \quad N \cap H = \{1_G\}.$$

When this occurs, G is said to be the (*internal*) *semidirect product of N by H*. Note that the First Isomorphism Theorem then shows that

$$G/N \cong H.$$

(i) Prove that every $a \in G$ is uniquely expressible as $a = nh$ with $n \in N$ and $h \in H$.

(ii) Since N is normal in G, there is a natural homomorphism $\psi\colon H \to Aut(N)$ given by conjugation, i.e.,

$$\psi(h)(n) = hnh^{-1}, \quad \text{for all } h \in H,\ n \in N.$$

Prove that for all $n, n' \in N$ and $h, h' \in H$

$$(nh)(n'h') = [n\,\psi(h)(n')]\,[hh']. \tag{2.34}$$

This formula shows that the multiplication in G is completely determined by the multiplication in N, the multiplication in H, and the homomorphism ψ.

2.65. Let G and L be groups and let $\alpha\colon G \to L$ be a homomorphism. Suppose there is a homomorphism $\beta\colon L \to G$ such that $\alpha \circ \beta = id_L$. Prove that G is the semidirect product of $ker(\alpha)$ by $im(\beta)$, with $im(\beta) \cong L$.

2.66. *External semidirect product.* The internal semidirect product motivates the definition of the external semidirect product: Let N and H be groups, and let $\theta\colon H \to Aut(N)$ be a group homomorphism. Let $S = N \times H$ as a set, with an operation $\cdot$ defined by

$$(n, h) \cdot (n', h') = (n\,\theta(h)(n'), hh') \tag{2.35}$$

for all $n, n' \in N$ and $h, h' \in H$. (Compare this formula with (2.34).)

(i) Prove that with this operation S is a group, with identity element $(1_N, 1_H)$ and that for all $n \in N$, $h \in H$,

$$(n, h)^{-1} = (\theta(h^{-1})(n^{-1}), h^{-1}).$$

The group S is called the (*external*) *semidirect product of N by H via θ*, and is denoted $N \rtimes_\theta H$.

(ii) In S, let

$$N' = \{(n, 1_H) \mid n \in N\} \quad \text{and} \quad H' = \{(1_N, h) \mid h \in H\}.$$

Prove that H' is a subgroup of S with $H' \cong H$ and N' is a subgroup of S with $N' \cong N$. Clearly, $S = N'H'$ and $N' \cap H' = 1_S$. Prove further that for any $n' = (n, 1_H) \in N'$ and $h' = (1_N, h) \in H'$,

$$h'n'h'^{-1} = (\theta(h)(n), 1_H) \in N'. \tag{2.36}$$

Thus, N' is normal in S and S is the internal semidirect product of N' and H' and the associated homomorphism $\psi\colon H' \to Aut(N')$ is given by

$$\psi\big((1_N, h)\big)\big((n, 1_H)\big) = (\theta(h)(n), 1_H).$$

That is, ψ corresponds to θ when we identify N' with N and H' with H. This shows that external semidirect products are "the same as" internal semidirect products, up to isomorphism.

(iii) Prove that

$$Z(N \rtimes_\psi H) = \left\{(n,h) \,\middle|\, \begin{array}{l} n \in Z(N) \text{ and } \psi(k)(n) = n \text{ for all } k \in H \\ \text{and } h \in Z(H) \cap ker(\psi) \end{array} \right\}.$$

Thus, $N \rtimes_\psi H$ is abelian iff N and H are abelian and ψ is the trivial homomorphism.

2.67. Let group G be the internal semidirect product of its subgroups N by H as in problem 2.64.

(i) Let K be a subgroup of G with $K \supseteq H$. Prove that K is the internal semidirect product of $K \cap N$ by H.

(ii) Prove that the map $K \mapsto K \cap N$ gives a one-to-one correspondence between the subgroups K of G with $K \supseteq H$, and the subgroups L of N satisfying $hLh^{-1} = L$ for all $h \in H$.

(iii) Let M be a subgroup of G with $M \supseteq N$. Prove that M is the internal semidirect of N by $M \cap H$.

(iv) Prove that the map $M \mapsto M \cap H$ gives a one-to-one correspondence between the subgroups of G containing N and the subgroups of H.

2.68. *Generalized dihedral groups.* This problem give a complete description of all generalized dihedral groups as semidirect products.

(i) Let G be a generalized dihedral group with distinguished subgroup A as in problem 2.12 above. Take any $b \in G \setminus A$. Prove that G is a semidirect product of A by $\langle b \rangle$ with $\langle b \rangle \cong C_2$.

(ii) Now, conversely take any abelian group A containing an element c with $|c| > 2$. Let $inv\colon A \to A$ be the inverse map given by $a \mapsto a^{-1}$ for all $a \in A$. Since A is abelian, $inv \in Aut(A)$, and $inv \neq id_A$ since $inv(c) \neq c$. Then, $\langle inv \rangle \cong C_2$, as $inv \circ inv = id_A$. Let ψ be the isomorphism $C_2 \to \langle inv \rangle$, viewed as a homomorphism $C_2 \to Aut(A)$. Let

$$D \;=\; A \rtimes_\psi C_2.$$

Prove that D is a generalized dihedral group with distinguished subgroup isomorphic to A.

2.69. The orthogonal group O_2 acts by left multiplication on the column space $\mathbb{R}^2$, and the action is $\mathbb{R}$-linear, hence compatible with the addition on $\mathbb{R}^2$. Thus, for the homomorphism $\tau\colon O_2 \to \Sigma(\mathbb{R}^2)$ associated to the group action as in problem 2.31(i), we have $im(\tau) \subseteq Aut(\mathbb{R}^2)$. Prove that

$$\mathbb{R}^2 \rtimes_\tau O_2 \;\cong\; \mathcal{RM},$$

where $\mathcal{RM}$ is the group of rigid motions on $\mathbb{R}^2$ as in problem 2.11. (This isomorphism corresponds to the internal semidirect decomposition of $\mathcal{RM}$ given in problem 2.11(v) as T by $\mathcal{RM}_0$.)

2.70. *Affine group of* $\mathbb{Z}_n$. Fix $n \in \mathbb{N}$. For any $a \in \mathbb{Z}_n$ and $u \in \mathbb{Z}_n^*$, the bijective function $f_{a,u}\colon \mathbb{Z}_n \to \mathbb{Z}_n$ given by $r \mapsto ur + a$ is called an *affine transformation* of $\mathbb{Z}_n$. Note that

$$f_{b,v} \circ f_{a,u} \;=\; f_{b+va,vu}\,, \quad \text{for all } a, b \in \mathbb{Z}_n \text{ and } u, v \in \mathbb{Z}_n^*. \tag{2.37}$$

The *affine group of* $\mathbb{Z}_n$ is

$$Aff_n \;=\; \{f_{a,u} \mid a \in \mathbb{Z},\ u \in \mathbb{Z}_n^*\}. \tag{2.38}$$

Note that Aff_n is a subgroup of the symmetric group $\Sigma(\mathbb{Z}_n)$. Also, (2.37) shows that the map $f_{a,u} \mapsto (a, u)$ gives an isomorphism

$$Aff_n \;\cong\; \mathbb{Z}_n \rtimes_\mu \mathbb{Z}_n^*, \tag{2.39}$$

where $\mu\colon \mathbb{Z}_n^* \to Aut(\mathbb{Z}_n)$ is the isomorphism given by $u \mapsto f_{0,u}$. We identify the symmetric group S_n with $\Sigma(\mathbb{Z}_n)$ via the one-to-one correspondence $\{1, 2, \ldots, n\} \leftrightarrow \mathbb{Z}_n$ given by $i \leftrightarrow [i]_n$. Prove that when Aff_n is thus viewed as a subgroup of S_n,

$$Aff_n \;=\; N_{S_n}\big(\langle (1\ \ 2\ \ \ldots\ \ n)\rangle\big), \tag{2.40}$$

the normalizer in S_n of its cyclic subgroup generated by the n-cycle $(1\ \ 2\ \ \ldots\ \ n)$. (Hint: Recall problem 2.39.)

The holomorph of a group. Let G be any group. The identity map id from $Aut(G)$ to itself is a group homomorphism. Set

$$Hol(G) \;=\; G \rtimes_{id} Aut(G). \tag{2.41}$$

$Hol(G)$ is called the *holomorph of* G.

Example 2.71.

(i) Since for the cyclic subgroup C_n we have an isomorphism $\psi\colon \mathbb{Z}_n^* \to Aut(C_n)$ (see (2.15)),

$$Hol(C_n) \;\cong\; C_n \rtimes_{\psi} \mathbb{Z}_n^*, \tag{2.42}$$

a group of order $n\varphi(n)$, which is nonabelian if $n \geq 3$ and isomorphic to C_2 if $n = 2$. Explicitly, if $C_n = \langle\omega\rangle$, then the multiplication in $Hol(C_n)$ is given by

$$(\omega^i, [k]_n) \cdot (\omega^j, [\ell]_n) \;=\; (\omega^{i+kj}, [k\ell]_n).$$

Clearly, also from (2.42) and (2.39),

$$Hol(C_n) \;\cong\; Aff_n. \tag{2.43}$$

(ii) Let p and q be primes with $q \equiv 1\ (mod\ p)$. Since

$$p \mid (q-1) \;=\; \varphi(q) \;=\; \big|\mathbb{Z}_q^*\big|,$$

by Cauchy's Theorem there is $[s]_q \in \mathbb{Z}_q^*$ of order p. Let $\rho\colon \langle [s]_q\rangle \to Aut(C_q)$ be the restriction of the map ψ of part (i), and let $G = C_q \rtimes_{\rho} \langle [s]_q\rangle$. Then G is a nonabelian group of order pq which is isomorphic to a subgroup of $Hol(C_q)$. Also, taking $a = (\omega, [1]_q)$ where $C_q = \langle\omega\rangle$, and $b = (1, [s]_q)$, we have

$$G \;=\; \langle a, b\rangle \quad \text{with} \quad a^q = 1, \quad b^p = 1, \quad \text{and} \quad bab^{-1} = a^s.$$

(iii) Let p be a prime number with $p > 2$. The binomial expansion shows that $(1+p)^p \equiv 1 \ (mod\ p^2)$ (recall (1.16)); hence, $[1+p]_{p^2}$ has order p in the group $\mathbb{Z}^*_{p^2}$. Let

$$\tau\colon \langle [1+p]_{p^2} \rangle \ \to \ Aut(C_{p^2})$$

be the restriction to $\langle [1+p]_{p^2} \rangle$ of the map ψ of part (i). Let

$$P \ = \ C_{p^2} \rtimes_\tau \langle [1+p]_{p^2} \rangle.$$

Then, P is a nonabelian group of order p^3 which is isomorphic to a subgroup of $Hol(C_{p^2})$. If $C_{p^2} = \langle \omega \rangle$, then, taking $c = (\omega, [1]_{p^2})$ and $d = (1, [1+p]_{p^2})$ in P, we have

$$P \ = \ \langle c, d \rangle \ \text{ with } \ c^{p^2} = 1, \ \ d^p = 1, \ \text{ and } \ dcd^{-1} = c^{1+p}.$$

2.72. Let $\mathcal{K}_4$ be the Klein 4-group as in (2.32) above. So, $\mathcal{K}_4$ is the normal subgroup of S_4 of order 4 and $\mathcal{K}_4 \cong C_2 \times C_2$. We view S_3 as a subgroup of S_4 by identifying it with $\{\sigma \in S_4 \mid \sigma(4) = 4\}$.

(i) Prove that S_4 is the internal semidirect product of $\mathcal{K}_4$ by S_3.

(ii) Prove that

$$Aut(\mathcal{K}_4) \ \cong \ S_3 \quad \text{and} \quad Hol(\mathcal{K}_4) \ \cong \ S_4. \tag{2.44}$$

2.73. *Automorphism groups of semidirect products.* This problem describes a situation in which the automorphism group of a semidirect product of groups is again a semidirect product. Let $G = NH$ be an internal semidirect product as in problem 2.64, with associated homomorphism $\theta\colon H \to Aut(N)$. Suppose that N is a *characteristic subgroup* of G, i.e., that every automorphism of G maps N onto itself; suppose further that $im(\theta) \subseteq Z(Aut(N))$.

(i) Take $\tau \in Aut(N)$, and define $\tau'\colon G \to G$ by

$$\tau'(nh) \ = \ \tau(n)h \qquad \text{for all } \ n \in N, \, h \in H.$$

Prove that $\tau' \in Aut(G)$. Prove further that the map $\beta\colon Aut(N) \to Aut(G)$ given by $\tau \mapsto \tau'$ is a group homomorphism.

(ii) Prove that there is a well-defined group homomorphism

$$\alpha\colon \ Aut(G) \longrightarrow Aut(N) \quad \text{given by } \ \sigma \mapsto \sigma|_N,$$

where $\sigma|_N$ is the the restriction of σ to N. Prove further that $\alpha \circ \beta = id_{Aut(N)}$. Deduce from problem 2.65 that $Aut(G)$ is the semidirect product of $ker(\alpha)$ by $im(\beta)$, with

$$ker(\alpha) \;=\; \{\sigma \in Aut(G) \mid \sigma|_N = id_N\}$$

and $im(\beta) \cong Aut(N)$.

2.74. Let G be a generalized dihedral group with distinguished subgroup A. Note that problem 2.28(iv) shows that A is a characteristic subgroup of G.

(i) Prove that $\{\sigma \in Aut(G) \mid \sigma|_A = id_A\} \cong A$.

(ii) Prove that $Aut(G) \cong Hol(A)$.

For example, since the dihedral group D_n for $n \geq 3$ has distinguished subgroup isomorphic to C_n,

$$Aut(D_n) \;\cong\; Hol(C_n) \;\cong C_n \rtimes_\psi \mathbb{Z}_n^*, \tag{2.45}$$

a group of order $n\varphi(n)$ (see Example 2.71(i) above).

2.75. Let p be a prime number, and let G be a group of order $p^a b$ where $p \nmid b$ and $a \in \mathbb{N}$. Suppose that G has a cyclic normal p-Sylow subgroup and a subgroup of order b. Prove that G and $Aut(G)$ are semidirect products of proper subgroups.

Wreath products. Let A and B be any groups. Let $B^A = \prod_{a \in A} B$, the direct product of $|A|$ copies of B, with its usual direct product group structure. (B^A can also be identified with the set of functions from A to B.) Define a group homomorphism $\theta\colon A \to Aut(B^A)$ by

$$\theta(c)(\ldots, \underset{a}{b_a}, \ldots) \;=\; (\ldots, \underset{a}{b_{c^{-1}a}}, \ldots).$$

That is, $\theta(c)$ acts on an element in B^A by permuting its components according to the left multiplication action of A on A. (For $\gamma \in B^A$, the ca-component of $\theta(c)(\gamma)$ is the a-component of γ.) The *wreath product* of B by A is defined to be the semidirect product

$$B \wr A \;=\; B^A \rtimes_\theta A. \tag{2.46}$$

Note that if A and B are finite, then $\left|B \wr A\right| = |A| \cdot |B|^{|A|}$.

2.76. *Sylow subgroups of symmetric groups.* Let p be a prime number.

(i) Let P be a p-Sylow subgroup of S_{p^2}. Prove that $P \cong C_p \wr C_p$.

(ii) Determine the number of p-Sylow subgroups of S_{p^2}.

(iii) More generally, define inductively

$$W_1 = C_p,\ W_2 = C_p \wr C_p,\ \ldots,\ W_i = W_{i-1} \wr C_p,\ \ldots . \quad (2.47)$$

Prove that for each $k \in \mathbb{N}$, every p-Sylow subgroup of S_{p^k} is isomophic to W_k.

Note: For $r \in \mathbb{R}$, let $[\![r]\!]$ denote the largest integer k such that $k \leq r$. Then, for $n \in \mathbb{N}$, the largest power ℓ of a prime p dividing $n!$ is $\ell = [\![\frac{n}{p}]\!] + [\![\frac{n}{p^2}]\!] + [\![\frac{n}{p^3}]\!] + \ldots$ (a finite sum). If we take the base-p expansion of n,

$$n = c_k p^k + c_{k-1} p^{k-1} + \ldots + c_i p^i + \ldots + c_1 p + c_0,$$

with each $c_i \in \{0, 1, 2, \ldots, p-1\}$, then $\ell = \sum_{i=1}^{k} c_i \big(\frac{p^i - 1}{p-1}\big)$. Using part (iii) of the preceding problem, one can show that every p-Sylow subgroup of S_n (which has order p^ℓ) is isomorphic to

$$\big(W_k\big)^{c_k} \times \big(W_{k-1}\big)^{c_{k-1}} \times \ldots \times \big(W_i\big)^{c_i} \times \ldots \times \big(W_1\big)^{c_1},$$

where $\big(W_i\big)^{c_i}$ denotes a direct product of c_i copies of the W_i of (2.47).

2.77. Let G be a finite nonabelian group in which every proper subgroup is abelian. Prove that G has a nontrivial proper normal subgroup.

2.78. For $n \in \mathbb{N}$ with $n \geq 2$, prove that every group of order n is cyclic iff $gcd(n, \varphi(n)) = 1$. (In view of the formula (1.11) for $\varphi(n)$, the condition on n is equivalent to: $n = p_1 p_2 \ldots p_k$ for distinct primes $p_1, \ldots, p_k$ satisfying $p_i \not\equiv 1 \ (mod\ p_j)$ for all i, j.) (Hint: Use the preceding problem.)

Cyclic factor groups. Let G be a group with a normal subgroup N such that G/N is cyclic of finite order s. Take any $a \in G$ such that $\langle aN \rangle = G/N$, and let $b = a^s \in N$. Let $\gamma_a \colon G \to G$ be the inner automorphism determined by a, mapping g in G to aga^{-1}. Thus, $\gamma_a(b) = b$ as $b = a^s$. Note that every $g \in G$ is expressible uniquely

as $g = a^i n$ for some integer i with $0 \leq i \leq s-1$ and $n \in N$. If for $g' \in G$ we express g' in the same form as $g' = a^j n'$, then

$$gg' = a^{i+j}(a^{-j}na^j)n' = \begin{cases} a^{i+j}[\gamma_a^{-j}(n)n'], & \text{if } i+j < s; \\ a^{i+j-s}[b\,\gamma_a^{-j}(n)n'], & \text{if } i+j \geq s. \end{cases}$$

Thus, the multiplication in G is entirely determined by data from N, namely the multiplication in N, the automorphism $\gamma_a|_N$ of N, and the element b of N satisfying $\gamma_a(b) = b$ and $(\gamma_a|_N)^s = \gamma_b$ in $\mathcal{I}nn(N)$. The next problem shows that these data are enough to determine all groups G containing N as a normal subgroup with G/N finite cyclic.

2.79. Let N be any group, and let $s \in \mathbb{N}$. Take any $b \in N$ and suppose there is $\alpha \in Aut(N)$ such that $\alpha^s = \gamma_b$, the conjugation by b automorphism. Let $C = \langle a \rangle$ be an infinite cyclic group, and let $\psi\colon C \to Aut(N)$ be the homomorphism given by $a^i \mapsto \alpha^i$ for $i \in \mathbb{Z}$. Consider the semidirect product $H = N \rtimes_\psi C$.

(i) Prove that $(b, a^{-s}) \in Z(H)$.

(ii) Let $K = \langle (b, a^{-s}) \rangle$, the cyclic subgroup of H generated by (b, a^{-s}), which is a normal subgroup of H since $K \subseteq Z(H)$ by part (i). Let $N_1 = \{(n, 1_C) \mid n \in N\}$, the usual isomorphic copy of N in H. Prove that $N_1 \cap K = \{1_H\}$.

(iii) Let $G = H/K$, and let $N_2 = N_1K/K$, the image of N_1 in G; so, $N_2 \cong N_1 \cong N$ by part (ii). Let a' be the image of $(1_N, a)$ in G and b' the image of $(b, 1_C)$ in N_2. Prove that N_2 is a normal subgroup of G, and G/N_2 is a cyclic group of order s with generator $a'N_2$; prove further that $a'^s = b'$ and conjugation by a' on N_2 corresponds to the automorphism α on N.

Recall (see problem 2.52) that every p-group has a normal subgroup of index p with cyclic factor group. Thus, the preceding problem shows in principle how all groups of order p^r can be constructed from those of order p^{r-1}.

2.8. Free groups and groups by generators and relations

Free groups. Let S be a set. A *free group on* S is a pair $(F(S), \iota)$, where $F(S)$ is a group, $\iota\colon S \to F(S)$ is a function, and the following universal mapping property holds: For every group G and every function $f\colon S \to G$, there is a unique group homomorphism $\alpha_f\colon F(S) \to G$ such that $\alpha_f \circ \iota = f$. For any set S it is known that there is a free group on S—see, e.g., Hungerford [**9**, p. 65], or Dummit & Foote [**5**, pp. 216–217] or Rotman [**20**, pp. 299–301]. For all the problems given here, the existence of $(F(S), \iota)$ is needed, but the explicit construction is not, and one can work entirely from the universal mapping property.

2.80. Let $(F(S), \iota)$ be a free group on a set S, and $(F(T), j)$ a free group on a set T. Suppose there is a bijection $f\colon S \to T$. Prove that for the function $j \circ f\colon S \to F(T)$ the associated group homomorphism $\alpha_{j\circ f}\colon F(S) \to F(T)$ is an isomorphism, with inverse $\alpha_{\iota\circ f^{-1}}\colon F(t) \to F(S)$. It follows, by taking $T = S$, that the group $F(S)$ is determined up to isomorphism by S, so $(F(S), \iota)$ is called "the" free group on S.

2.81. Let $(F(S), \iota)$ be a free group on a set S.

(i) Prove that the map $\iota\colon S \to F(S)$ is injective. Because of this, it is customary to view S as a subset of $F(S)$ by identifying S with $\iota(S)$. We then write $F(S)$ instead of $(F(S), \iota)$ for the free group on S.

(ii) Prove that $F(S) = \langle S \rangle$.

Let G be a group, and T a subset of G. The *normal subgroup of G generated by T* is $\bigcap N$ where N ranges over the normal subgroups of G containing T. It is easy to check that this group equals $\langle \{gTg^{-1} \mid g \in G\} \rangle$.

2.82. Let $F(S)$ be the free group on a set S. Let T be a subset of S, and let N be the normal subgroup of $F(S)$ generated by T. Prove that $F(S)/N \cong F(R)$ for some free group $(F(R), j)$.

Generators and relations. Let $\{a_i\}_{i\in I}$ be a subset of a group G. A *word* $w(a_i)$ in the a_i is a specified product $t_1t_2\dots t_n$ in G, where each $t_j \in \{a_i\}_{i\in I} \cup \{a_i^{-1}\}_{i\in I}$. If $\{c_i\}_{i\in I}$ is a subset of a group H, then $w(c_i)$ denotes the corresponding word in the c_i, i.e., the product in H where each appearance of a_i (resp. a_i^{-1}) in $w(a_i)$ is replaced by c_i (resp. c_i^{-1}). A *relation* among the a_i is an equation of the form $w(a_i) = w'(a_i)$, where $w(a_i)$ and $w'(a_i)$ are words in the a_i. Since $w(a_i) = w'(a_i)$ is equivalent to $w(a_i)w'(a_i)^{-1} = 1_G$, it suffices to consider relations of the form $w(a_i) = 1_G$.

The *group with generators* a_i *for* $i \in I$ *and relations* $w_j(a_i) = 1$ *for* $j \in J$ (where each $w_j(a_i)$ is a word in the a_i), denoted

$$\langle a_i,\ i \in I \mid w_j(a_i) = 1,\ j \in J\rangle, \tag{2.48}$$

is defined to be the group $F(S)/R$, where $S = \{s_i\}_{i\in I}$ with the s_i distinct for distinct i, R is the normal subgroup of $F(S)$ generated by the $w_j(s_i)$ for all $j \in J$, and each $a_i = s_iR$ in $F(S)/R$. Note that the a_i generate $F(S)/R$, as $F(S) = \langle S\rangle$, and since each $w_j(s_i) \in R$, we have $w_j(a_i) = 1_{F(S)/R}$.

Let $G = \langle a_i,\ i \in I \mid w_j(a_i) = 1,\ j \in J\rangle$. The universal mapping property for G is: Let H be a group and let $\{b_i\}_{i\in I}$ be a subset of H. If $w_j(b_i) = 1_H$ for each $j \in J$, then there is a unique group homomorphism $\beta\colon G \to H$ such that $\beta(a_i) = b_i$ for each $i \in I$. (Proof: Write $G = F(S)/R$ as above. There is a unique homomorphism $\alpha\colon F(S) \to H$ with $\alpha(s_i) = b_i$ for all $i \in I$. Since $w_j(b_i) = 1_H$, we have $w_j(s_i) \in ker(\alpha)$ for every $j \in J$. So, $R \subseteq ker(\alpha)$ as $ker(\alpha) \triangleleft F(S)$. Hence, by the FHT there is a group homomorphism β from $G = F(S)/R$ to H such that $\beta(a_i) = \beta(s_iR) = \alpha(s_i) = b_i$. Since $G = \langle\{a_i\}_{i\in I}\rangle$, the homomorphism β on G is completely determined by the $\beta(a_i)$.)

Example 2.83. For $n \geq 3$, we prove that

$$D_n \cong \langle a, b \mid a^n = 1,\ b^2 = 1,\ \text{and } (ba)^2 = 1\rangle.$$

Let $G = \langle a, b \mid a^n = 1,\ b^2 = 1,\ \text{and } (ba)^2 = 1\rangle$. Now, D_n contains the rotation ρ about the origin through the angle $\frac{2\pi}{n}$, and also contains a reflection τ about an axis of symmetry of a regular n-gon. Note that in D_n, $\rho^n = 1$, $\tau^2 = 1$, and, as $\tau\rho$ is a reflection, $(\tau\rho)^2 = 1$.

By the universal mapping property for G, there is a homomorphism $\beta\colon G \to D_n$ with $\beta(a) = \rho$ and $\beta(b) = \tau$. Then,

$$im(\beta) \supseteq \langle \beta(a), \beta(b) \rangle = \langle \rho, \tau \rangle = D_n.$$

Hence, β is surjective and $|G| \geq |\, im(\beta)| = 2n$. Let $N = \langle a \rangle \subseteq G$; so $|N| = |a| \leq n$, as $a^n = 1$. Since $(ba)^2 = 1$ and $b^{-1} = b$, we have $bab^{-1} = a^{-1}$. Hence, b lies in the normalizer $N_G(N)$. Since in addition, $a \in N \subseteq N_G(N)$, we have $G = \langle a, b \rangle \subseteq N_G(N)$; so, $N \lhd G$. Note that $G/N = \langle aN, bN \rangle = \langle bN \rangle$, as $aN = 1_{G/N}$. Hence,

$$|G/N| = |bN| \leq |b| \leq 2,$$

as $b^2 = 1$. Thus, by Lagrange's Theorem

$$|G| = |G/N| \cdot |N| \leq 2n = |D_n|.$$

Since $\beta\colon G \to D_n$ is surjective and $|G| \leq |D_n|$, the map β must be bijective: hence β is an isomorphism, completing the proof.

Note that from the presentation of G we proved that $|G| \leq 2n$. However, without the map β of G onto the known group D_n, it would be extremely difficult to show that $|G| = 2n$. For while the relations show that $G = \{1, a, a^2, \ldots, a^{n-1}, b, ba, ba^2, \ldots, ba^{n-1}\}$ it is hard to rule out the possibility that the relations might imply that some of the listed elements of G are the same. The convincing way to verify that these elements are all different is to observe that their images in D_n are all different.

2.84. Let p and q be prime numbers with $q \equiv 1 \ (mod\ p)$.

(i) Since $p|(q-1) = |\mathbb{Z}_q^*|$, by Cauchy's Theorem there is an element $[s]_q$ of order p in the group $\mathbb{Z}_q^*$. Let

$$G = \langle a, b \mid a^q = 1,\ b^p = 1,\ bab^{-1} = a^s \rangle.$$

Prove that group G is nonabelian of order pq. (Hint: Recall Example 2.71(ii).)

(ii) Let H be a nonabelian group of order pq. Prove that H is isomorphic to the group G of part (i). You may use the fact that $\mathbb{Z}_q^*$ is a cyclic group; see problem 3.32 below.

Note: For prime numbers p and q with $p < q$, if C is an abelian group of order pq, then C is cyclic, since if we take $c, d \in C$ with $o(c) = p$ and $o(d) = q$, then $o(cd) = pq$. If $p \nmid (q-1)$, let K be any group with $|K| = pq$. Then, the Sylow theorems show that each Sylow subgroup of K is normal in K, from which it follows that K is the direct product of its (abelian) Sylow subgroups. Hence, K is abelian, so $K \cong C$. Thus, when $p \nmid (q-1)$ there is up to isomorphism only one group of order pq. But when $p|(q-1)$, there are two isomorphism classes: that of the cyclic group and that of the group of the preceding problem.

2.85. For any integer $n \geq 2$, let Q_n be the generalized quaternion group of order $4n$, as in problem 2.9.

(i) Prove that

$$Q_n \cong \langle a, b \mid a^{2n} = 1,\ b^2 = a^n,\ bab^{-1} = a^{-1} \rangle. \tag{2.49}$$

(ii) If $n \geq 3$, prove that

$$Aut(Q_n) \cong Hol(C_{2n}), \tag{2.50}$$

a group of order $2n\varphi(2n)$.

(iii) Prove that

$$Aut(Q_2) \cong Hol(\mathcal{K}_4) \cong S_4, \tag{2.51}$$

where $\mathcal{K}_4$ is the Klein 4-group (cf. problem 2.72).

2.86. Let $G = \langle a, b \mid ab = ba \rangle$. Prove that $G \cong \mathbb{Z} \times \mathbb{Z}$.

2.87. Prove that for any prime p, the group

$$\langle a, b \mid a^{p^2} = 1,\ b^p = 1,\ bab^{-1} = a^{1+p} \rangle$$

is nonabelian of order p^3. (See Example 2.71(iii).)

2.88. For any prime p, let G be a nonabelian group of order p^3 containing an element of order p^2.

(i) Prove that if $p \neq 2$, then

$$G \cong \langle a, b \mid a^{p^2} = 1,\ b^p = 1,\ bab^{-1} = a^{1+p} \rangle. \tag{2.52}$$

For this, the following identity is useful: For any c, d in a group and any $s \in \mathbb{N}$, if $dcd^{-1} = c^s$, then for $k \in \mathbb{N}$,

$$\begin{aligned}(cd)^k &= c(dcd^{-1})(d^2cd^{-2})\dots(d^icd^{-i})\dots(d^{k-1}cd^{-(k-1)})\,d^k \\ &= c^t d^k,\end{aligned}$$

where

$$t = 1 + s + s^2 + \dots + s^{k-1} = (s^k - 1)/(s-1).$$

(ii) If $p \neq 2$, determine $|Aut(G)|$.

(iii) If $p = 2$, prove that $G \cong D_4$ or $G \cong Q_2$. (Their automorphism groups are described in (2.45) and (2.51).)

2.89. Let p be a prime number, and let $N = \langle c\rangle \times \langle d\rangle$ where $|c| = |d| = p$; so $N \cong C_p \times C_p$.

(i) Define $\beta\colon N \to N$ by

$$\beta(c^i d^j) = c^{i+j}d^j \qquad \text{for all } \ i, j \in \mathbb{Z}.$$

Prove that β is well-defined and that $\beta \in Aut(N)$ with $|\beta| = p$.

(ii) Let $\iota\colon \langle\beta\rangle \to Aut(N)$ be the inclusion map given by $\beta^i \mapsto \beta^i$ for all $i \in \mathbb{Z}$, and let $P = N \rtimes_\iota \langle\beta\rangle$. Prove that P is a nonabelian group of order p^3, and if $p \geq 3$, then every nonidentity element of P has order p. (When $p = 2$, this is not true, since then $P \cong D_4$.)

2.90. Let p be a prime number with $p \geq 3$, and let

$$G = \langle\, a, b, c \mid a^p = b^p = c^p = 1,\ ba = ab,\ ca = ac,\ cb = abc \,\rangle.$$

(i) Prove that for the Heisenberg group $\mathcal{H}(\mathbb{Z}_p)$ of problem 2.58, $\mathcal{H}(\mathbb{Z}_p) \cong G$.

(ii) Let P be the group of problem 2.89(ii). Prove that $P \cong G$.

2.91. For any integer $n \geq 4$, prove that

$$S_n \cong \Big\langle\, a_1, a_2, \dots, a_{n-1} \;\Big|\; \begin{array}{l} a_i^2 = 1,\ (a_ia_j)^3 = 1, \text{and} \\ (a_ia_ja_ia_k)^2 = 1 \text{ for all distinct } i, j, k \end{array} \Big\rangle.$$

(Hint: Consider the generators $(1\ 2), (1\ 3), \dots, (1\ n)$ of S_n.)

2.9. Nilpotent, solvable, and simple groups

Commutators. For any elements a, b of a group G, the *commutator of a and b* is

$$[a, b] \;=\; aba^{-1}b^{-1}.$$

Thus, $[a, b] = 1_G$ iff $ab = ba$. For subgroups H, K of G, define

$$[H, K] \;=\; \langle\{[h, k] \mid h \in H] \text{ and } k \in K\}\rangle. \tag{2.53}$$

Note that since

$$g[h, k]g^{-1} \;=\; [ghg^{-1}, gkg^{-1}] \quad \text{for all } g \in G,\ h \in H,\ k \in K,$$

if H and K are normal subgroups of G, then $[H, K]$ is also normal in G. In particular, the *derived group* (or *commutator subgroup*) of G is

$$G' \;=\; [G, G] \;=\langle\{aba^{-1}b^{-1} \mid a, b \in G\}\rangle. \tag{2.54}$$

Then, G' is a normal subgroup of G and G/G' is abelian. Indeed, note that for any normal subgroup K of G, the factor group G/K is abelian iff $K \supseteq G'$.

A group G is said to be *nilpotent* if it has normal subgroups $K_0, K_1, \ldots, K_k$ such that

$$\begin{aligned} &G \;=\; K_0 \;\supseteq K_1 \;\supseteq\; \ldots \supseteq K_k \;=\{1_G\},\\ &\qquad \text{with}\ \ K_{i-1}/K_i \subseteq Z(G/K_i) \quad \text{for} \quad i \in \{1, 2, \ldots, k\}. \end{aligned} \tag{2.55}$$

Note that

$$K_{i-1}/K_i \;\subseteq\; Z(G/K_i) \ \text{ is equivalent to } \ [G, K_{i-1}] \;\subseteq\; K_i. \tag{2.56}$$

Nilpotence of G is also characterized in terms of its *descending central series*; this is the chain of normal subgroups

$$G_{(0)} \;\supseteq\; G_{(1)} \;\supseteq\; G_{(2)} \;\supseteq\; \ldots \;\supseteq\; G_{(i)} \;\supseteq\; \ldots\,,$$

where

$$G_{(0)} \;=\; G,\ G_{(1)} \;=\; [G, G_{(0)}],\ \ldots,\ G_{(i)} \;=\; [G, G_{(i-1)}],\ \ldots\,.$$

Then, G is nilpotent iff $G_{(k)} = \{1_G\}$ for some k. (Proof: If $G_{(k)} = \{1_G\}$, set $K_i = G_{(i)}$ and note that (2.55) holds for these K_i by (2.56). Conversely, if we have K_i as in (2.55), then each $G_{(i)} \subseteq K_i$ by induction on i using (2.56). Hence, $G_{(k)} \subseteq K_k = \{1_G\}$.)

2.92. Let G be a finite group. Prove that the following conditions are equivalent:

(a) G is nilpotent.

(b) For every proper normal subgroup K of G, the group $Z(G/K)$ is nontrivial.

(c) For every proper subgroup H of G, $N_G(H) \supsetneq H$.

(d) Every maximal subgroup of G is a normal subgroup of G.

(e) Every Sylow subgroup of G is a normal subgroup of G.

(f) G is a direct product of p-groups (so a direct product of its Sylow subgroups).

2.93. Let G be a finite group. Prove that its Frattini subgroup $\mathcal{D}(G)$ is a nilpotent group. (Hint: Use problem 2.63.)

2.94. Let G be a nilpotent group, and let S be a subset of G. Prove that if $G = \langle \{gsg^{-1} \mid s \in S,\ g \in G\} \rangle$, then $G = \langle S \rangle$.

Solvable groups. A group G is said to be *solvable* if there is a chain of subgroups

$$G_0 = G \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_k = \{1_G\}$$

such that each G_{i+1} is a normal subgroup of G_i with G_i/G_{i+1} abelian. Note that if G is solvable, then every subgroup of G is solvable. Moreover, if K is a normal subgroup of G, then G is solvable iff K and G/K are each solvable. The proofs of these facts are not difficult and can be found in any text. Abelian groups, p-groups, and nilpotent groups are all solvable. Solvability of G is also characterized in terms of the *derived series of* G: This is the descending chain of normal subgroups

$$G^{(0)} \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \ldots \supseteq G^{(i)} \supseteq \ldots,$$

where

$$G^{(0)} = G,\ G^{(1)} = [G, G],\ G^{(2)} = [G^{(1)}, G^{(1)}],\ \ldots,$$
$$G^{(i)} = [G^{(i-1)}, G^{(i-1)}],\ \ldots.$$

(So $G^{(i)}/G^{(i-1)}$ is abelian for each i.) Then, G is solvable iff $G^{(k)} = \{1_G\}$ for some k.

2.95. Let G be a finite solvable group.

(i) Let M be a maximal proper normal subgroup of G. Prove that $G/M \cong C_p$ for some prime p.

(ii) Let K be a minimal nontrivial normal subgroup of G. Prove that K is an elementary abelian p-group, for some prime p.

(iii) Let H be a maximal (proper) subgroup of G. Prove that $|G{:}H|$ is a power of some prime number.

(iv) This example shows that the prime power in part (iii) above can be arbitrarily large: Let p be a prime number, and n any positive integer; let $\mathbb{F}$ be the finite field with $|\mathbb{F}| = p^n$, so $\mathbb{F}^* = \mathbb{F} \setminus \{0\}$. There is a homomorphism $\lambda\colon \mathbb{F}^* \to Aut(\mathbb{F})$ given by $\lambda(b)(c) = bc$ for all $b \in \mathbb{F}^*$ and $c \in \mathbb{F}$. Let $G = \mathbb{F} \rtimes_\lambda \mathbb{F}^*$, which is solvable since $\mathbb{F}$ and $\mathbb{F}^*$ are abelian, so solvable. Prove that the standard copy of $\mathbb{F}^*$ in G is a maximal subgroup of index p^n.

Triangular matrix groups. Let F be a field, and fix $n \in \mathbb{N}$ with $n \geq 2$. In the $n \times n$ matrix ring $M_n(F)$, for $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$ and $a \in F$, let $E_{ij}(a)$ be the matrix with ij-entry a, each kk-entry 1, and all other entries 0. That is,

$$E_{ij}(a) \;=\; I_n + ae_{ij}, \tag{2.57}$$

where I_n is the identity matrix and e_{ij} is the matrix with ij-entry 1 and all other entries 0. Such an $E_{ij}(a)$ is called an *elementary matrix*. Note that $E_{ij}(a)$ is a triangular matrix with $E_{ij}(a)^{-1} = E_{ij}(-a)$ and with determinant $det(E_{ij}(a)) = 1$. So, $E_{ij}(a) \in SL_n(F)$. Straightforward calculations yield the the following commutator identities:

$$\begin{aligned} [E_{ij}(a), E_{jk}(b)] \;&=\; E_{ik}(ab) \;\text{ if } k \neq i; \\ [E_{ij}(a), E_{ki}(b)] \;&=\; E_{kj}(-ab) \;\text{ if } k \neq j; \\ [E_{ij}(a), E_{k\ell}(b)] \;&=\; I_n \;\text{ if } j \neq k \text{ and } i \neq \ell. \end{aligned} \tag{2.58}$$

For the next problem, consider the the following sets of upper triangular matrices:

$$B = \left\{ C = (c_{ij}) \in M_n(F) \;\middle|\; \begin{array}{l} c_{kk} \in F^* \text{ for each } k \\ \text{and } c_{ij} = 0 \text{ for } i > j \end{array} \right\};$$

$$T = \left\{ C = (c_{ij}) \in M_n(F) \;\middle|\; \begin{array}{l} c_{kk} \in F^* \text{ for each } k \\ \text{and } c_{ij} = 0 \text{ for } i \neq j \end{array} \right\};$$

$$U = \left\{ C = (c_{ij}) \in M_n(F) \;\middle|\; \begin{array}{l} c_{kk} = 1 \text{ for each } k \\ \text{and } c_{ij} = 0 \text{ for } i > j \end{array} \right\};$$

$$U_\ell = \left\{ C = (c_{ij}) \in M_n(F) \;\middle|\; \begin{array}{l} c_{kk} = 1 \text{ for each } k \\ \text{and } c_{ij} = 0 \text{ for } i > j \\ \text{and for } i < j < i + \ell \end{array} \right\} \text{ for } \ell \in \mathbb{N}.$$

Note that B consists of all upper triangular matrices in $GL_n(F)$; T consists of the diagonal matrices in $GL_n(F)$; $U = U_1$ consists of all upper triangular matrices in $GL_n(F)$ with all 1's on the main diagonal (called unipotent matrices); and U_ℓ for $\ell \geq 2$ consists of the matrices in U with $\ell - 1$ diagonals of 0's just above the main diagonal. Thus, $U = U_1 \supseteq U_2 \supseteq \ldots$, with $U_m = \{I_n\}$ for $m \geq n$. Observe also that B, T, U, U_2, U_3, ... are all subgroups of $GL_n(F)$. Moreover, the process for row reducing a matrix in U to reach I_n shows that

$$U = \langle \{ E_{ij}(a) \mid i < j \text{ and } a \in F \} \rangle.$$

Likewise,

$$U_\ell = \langle \{ E_{ij}(a) \mid i + \ell < j \text{ and } a \in F \} \rangle.$$

Note also that $B = UT$ and $U \cap T = \{I_n\}$.

2.96. Let B, T, U, U_2, U_3, ... be the subgroups of $GL_n(F)$ just defined, for $n \geq 2$.

(i) Prove that $U, U_2, U_3, \ldots$ are each normal subgroups of B, with

$$B/U \cong T \cong \prod_{i=1}^{n} F^* \quad \text{and} \quad U_\ell / U_{\ell+1} \cong \prod_{i=1}^{n-\ell} F$$

for $\ell = 1, 2, \ldots, n-1$. Since U is normal in B, it follows that B is the semidirect product of U by T.

(ii) Prove that $U_\ell = [T, U_\ell] = [B, U_\ell]$ for all ℓ.

(iii) Prove that $[B, B] = U$.

(iv) Prove that $[U_\ell, U_m] = U_{\ell+m}$ for all ℓ, m.

(v) Deduce that U is nilpotent, and B is solvable but not nilpotent.

2.97. *Burnside's p^aq^b Theorem* says that for any distinct primes p and q and any $a, b \in \mathbb{N}$, every group of order p^aq^b is solvable. This is a difficult result, whose standard proof uses group representation theory (see, e.g., Dummit & Foote [**5**, pp. 886–890]). The special case considered here, when $b = 1$, is easier.

(i) Let G be any finite group, and let p be a prime number dividing $|G|$ such that G has more than one p-Sylow subgroup. Let E be a *maximal p-Sylow intersection* in G, i.e., $E = P_1 \cap P_2$ for distinct p-Sylow subgroups of P_1 and P_2 of G, and no subgroup of G properly containing E is expressible as such an intersection. Prove that there is a (well-defined!) one-to-one correspondence between (all) the p-Sylow subgroups of G containing E and (all) the p-Sylow subgroups of $N_G(E)$ given by $P \longleftrightarrow P \cap N_G(E)$.

(ii) Prove that every group of order p^aq is solvable, for any primes p and q, and any $a \in \mathbb{N}$.

2.98. Let p be a prime number, and let G be a solvable subgroup of S_p with $p \,\big|\, |G|$.

(i) Prove that G has a normal (hence unique) p-Sylow subgroup. (Hint: Apply problems 2.95(ii) and 2.37.)

(ii) Deduce that G is isomorphic to a subgroup of Aff_p. (Hint: Apply problem 2.70.)

(iii) Prove that for any $\sigma \in G$ with $\sigma \neq 1_G$,

$$\big|\{i \in \{1, 2, \ldots, p\} \mid \sigma(i) = i\}\big| \;=\; 1.$$

Simple groups. A group G is said to be *simple* if G is nontrivial and its only normal subgroups are G and $\{1_G\}$. Clearly, the only abelian simple groups are the cyclic groups C_p for p prime. It was noted above on p. 40 that the alternating groups A_n are simple for $n = 3$ and $n \geq 5$. We will see in problem 2.103 below that the $PSL_n(F)$ groups are nearly always simple.

2.99. Prove that the following assertions are equivalent:

(a) Every finite group of odd order is solvable.

(b) Every finite nonabelian simple group has even order.

Note: In fact, conditions (a) and (b) are both true! This is the amazing Feit–Thompson Theorem.

2.100.

(i) Let G be a finite nonabelian simple group with a proper subgroup H, and let $n = |G{:}H|$. Prove that $n \geq 5$ and that either $G \cong A_n$ or $G \cong A_{n-1}$ or $|G| < (n-1)!/2$.

(ii) Deduce that if a simple group G has order 60, then $G \cong A_5$.

(iii) Deduce also that there is no simple group of order 90 or 120.

(iv) Prove that S_n is not isomorphic to a subgroup of A_{n+1}, for every $n \geq 2$.

2.101. Let G be a nonabelian simple group. Prove that

$$Aut(Aut(G)) \cong Aut(G).$$

2.102. Let G be a group which acts on a set S, and let K be the kernel of the action. Suppose the following three conditions hold:

(i) The action is *doubly transitive*, i.e., for every $s, t \in S$ with $s \neq t$ and every $x, y \in S$ there is $g \in G$ such that both $g \cdot s = x$ and $g \cdot t = y$.

(ii) $G = [G, G]$.

(iii) For some $s \in S$ there is an abelian subgroup A of the stabilizer group G_s such that $A \lhd G_s$, but G is generated by all the conjugates gAg^{-1} for $g \in G$.

Prove that G/K is a simple group. (Hint: Take any normal subgroup L of G with $K \subsetneqq L$. Prove that $G = LG_s$; then, $LA \lhd G$; then $LA = G$; then, $L = G$.)

Simplicity of PSL groups. Let F be a field, and let $n \in \mathbb{N}$ with $n \geq 2$. Recall that $SL_n(F) = \{A \in M_n(F) \mid det(A) = 1\}$. It is easy to check that

$$Z(SL_n(F)) = \{dI_n \mid d \in F^* \text{ and } d^n = 1\}.$$

The *projective special linear group* of degree n for F is

$$PSL_n(F) \;=\; SL_n(F)/Z(SL_n(F)). \tag{2.59}$$

The next problem will show that PSL groups are nearly always simple. This will be done by applying the preceding problem with $G = SL_n(F)$ and $K = Z(SL_n(F))$. We will need the property that

$$SL_n(F) \;=\; \langle \{E_{ij}(a) \mid i \neq j,\ a \in F\} \rangle, \tag{2.60}$$

where the $E_{ij}(a)$ are the elementary matrices in $M_n(F)$, as in (2.57). This property is provable using the fact that the row reduction process for bringing a nonsingular matrix to diagonal form shows that every matrix C in $GL_n(F)$ is expressible as a product of elementary matrices E_i and a diagonal matrix D and permutation matrices P of transpositions in S_n. (For $\sigma \in S_n$, the associated permutation matrix $P_\sigma \in M_n(F)$ has each $\sigma(i)i$ entry 1 and all other entries 0.) The process still works if we use instead of P a modified permutation matrix P' with one of the 1-entries of P replaced by -1. (So, $det(P') = 1$.) If $C \in SL_n(F)$ then the resulting D satisfies $det(D) = 1$. Then, short calculation shows that each P' and D are products of elementary matrices. For example, with $n = 2$,

$$\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \;=\; E_{21}(1)E_{12}(-1)E_{21}(1),$$

and for $a \in F^*$ with $a \neq 1$,

$$\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right) \;=\; E_{12}(a^2 - a)E_{21}(a^{-1})E_{12}(1 - a)E_{21}(-1).$$

2.103. Let F be any field. Consider $SL_n(F)$ and $PSL_n(F)$ for $n \geq 2$.

(i) Prove that $SL_n(F) = [SL_n(F), SL_n(F)]$ except when $n = 2$ and $|F| = 2$ or 3. (Use (2.60) and the commutator identities (2.58).)

(ii) $SL_n(F)$ acts by left multiplication on the n-dimensional F-column space $F^n = \left\{ \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \;\middle|\; a_1, \ldots, a_n \in F \right\}$. Since the action is F-linear this yields an induced group action of $SL_n(F)$ on the set S of 1-dimensional F-vector subspaces of F^n. Prove that this action on S is doubly transitive, and that the kernel of the action is $Z(SL_n(F))$.

(iii) Let $s = F\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} \in S$. Determine the stabilizer group $SL_n(F)_s$.

(iv) Let

$$\begin{aligned} A &= \left\{ C = (c_{ij}) \in SL_n(F) \;\middle|\; \begin{array}{l} c_{ii} = 1 \text{ for all } i, \text{ and} \\ c_{ij} = 0 \text{ if } i > j \text{ or } 1 < i < j \end{array} \right\} \\ &= \langle \{ E_{1j}(a) \mid j = 2, 3, \ldots, n,\ a \in F \} \rangle \\ &= \left\{ \begin{pmatrix} 1 & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \;\middle|\; a_{12}, \ldots, a_{1n} \in F \right\}. \end{aligned}$$

Prove that A is an abelian normal subgroup of $SL_n(F)_s$ and that $SL_n(F) = \langle \{ gbg^{-1} \mid g \in SL_n(F),\ b \in A \} \rangle$. Thus, by problem 2.102, $PSL_n(F)$ is a simple group except when $n = 2$ and $|F| = 2$ or 3.

The preceding problem applies for both finite and infinite fields F. When $|F|$ is finite, we can compute the orders of the simple PSL groups as follows: Say $|F| = q < \infty$. It is known that q can be any power of any prime (see §5.7 below). Then,

$$|\, GL_n(F)| \;=\; (q^n - 1)\,(q^n - q)\,\ldots\,(q^n - q^i)\;\ldots\,(q^n - q^{n-1}). \quad (2.61)$$

To see this, recall that a matrix C in $M_n(F)$ is invertible iff its columns are linearly independent over F. Thus, for C to be in $GL_n(F)$ there are $q^n - 1$ choices for its first column, after which there are $q^n - q$ choices for the second column, since it cannot be in the F-linear span of the first column, etc. Since $SL_n(F)$ is the kernel of the surjective determinant homomorphism $GL_n(F) \to F^*$,

$$|\, SL_n(F)| \;=\; |\, GL_n(F)|/(q-1).$$

Also,

$$\big|Z(SL_n(F))\big| \;=\; \big|\{\, c \in F^* \mid c^n = 1 \,\}\big| \;=\; gcd(n, q-1),$$

since F^* is a cyclic group (see problem 3.32). Thus,

$$\big|PSL_n(F)\big| \;=\; \prod_{i=0}^{n-1} (q^n - q^i) \,/\, [(q-1)\, gcd(n, q-1)]. \quad (2.62)$$

2.10. Finite abelian groups

For an abelian group G, the *torsion subgroup of* G is

$$t(G) \;=\; \{a \in G \mid a^n = 1_G \text{ for some } n \in \mathbb{N}\}. \tag{2.63}$$

It is a subgroup of G as G is abelian. G is said to be a *torsion group* if $G = t(G)$. At the other extreme, G is *torsion-free* if $t(G) = \{1_G\}$. Note that if G is any abelian group, then $G/\,t(G)$ is torsion-free.

2.104. Prove that $t(\mathbb{R}/\mathbb{Z}) = \mathbb{Q}/\mathbb{Z}$.

For an abelian group G and any prime number p, the *p-primary component of* G is

$$G_{(p)} \;=\; \{a \in G \mid a^{p^n} = 1_G \text{ for some } n \in \mathbb{N}\}, \tag{2.64}$$

which is a subgroup of G.

2.105. Prove that for every prime p,

$$(\mathbb{Q}/\mathbb{Z})_{(p)} \;=\; \mathbb{Z}[\tfrac{1}{p}]/\mathbb{Z},$$

where

$$\mathbb{Z}[\tfrac{1}{p}] \;=\; \bigcup_{i=1}^{\infty} \tfrac{1}{p^i}\mathbb{Z} \;=\; \{\tfrac{r}{p^i} \mid r \in \mathbb{Z} \text{ and } i \in \mathbb{N}\}.$$

2.106. *Primary decomposition.* Let G be a finite abelian group, and let $p_1, p_2, \ldots, p_k$ be the distinct prime divisors of $|G|$. Prove that

$$G \;=\; G_{(p_1)} \times G_{(p_2)} \times \ldots \times G_{(p_k)}. \tag{2.65}$$

This is called the *primary decomposition* of G.

Example 2.107. For a cyclic group the primary decomposition is given by the Chinese Remainder Theorem (see (2.12)): For any distinct primes $p_1,\ldots,p_k$ and any $r_i \in \mathbb{N}$,

$$C_{p_1^{r_1}\ldots p_k^{r_k}} \;\cong\; C_{p_1^{r_1}} \times \ldots \times C_{p_i^{r_i}} \times \ldots \times C_{p_k^{r_k}}. \tag{2.66}$$

2.108. Let G be any torsion abelian group. Prove the *primary decomposition of* G:

$$G \;=\; \bigoplus_{p \text{ prime}} G_{(p)}. \tag{2.67}$$

The direct sum is over all the prime numbers p.

2.109. For a finite abelian group G the *exponent* of G is defined to be

$$\begin{aligned} exp(G) &= lcm\{\,|a| \;\big|\; a \in G\,\} \\ &= min\{\,n \in \mathbb{N} \mid a^n = 1_G \text{ for all } a \in G\,\}. \end{aligned} \tag{2.68}$$

Note that $exp(G) \,\big|\, |G|$, by Lagrange's Theorem.

(i) Prove that there is $a \in G$ with $|a| = exp(G)$.

(ii) Deduce that G is cyclic iff $exp(G) = |G|$.

The *Fundamental Theorem for Finite Abelian Groups* says: For every nontrivial finite abelian group A, we have

$$A \cong C_{n_1} \times C_{n_2} \times \ldots \times C_{n_k}, \tag{2.69}$$

where C_{n_i} is a cyclic group of order n_i, with $n_2|n_1$, $n_3|n_2$, ..., $n_i|n_{i-1}$, ..., $n_k|n_{k-1}$, and $n_k > 1$. Moreover, the positive integers $n_1, n_2, \ldots, n_k$ (subject to the divisibility conditions) are uniquely determined by A. They are called the *invariant factors* of A.

The next sequence of problems develops the notion of duality for finite abelian groups, and shows how it can be used to prove the Fundamental Theorem. Thus, the problems are intended to be solved without invoking the Fundamental Theorem.

Hom groups. For groups G, A with A abelian, let

$$Hom(G, A) = \{\text{the set of group homomorphisms } G \to A\}. \tag{2.70}$$

Define an operation $\cdot$ on $Hom(G, A)$ by $(f_1 \cdot f_2)(g) = f_1(g)f_2(g)$ for all $f_1, f_2 \in Hom(G, A)$ and $g \in G$. Note that $f_1 \cdot f_2$ is a group homomorphism as A is abelian. It is easy to check that with this operation $Hom(G, A)$ is an abelian group with identity element the constant function $\mathbf{1}\colon G \to A$ sending each $g \in G$ to 1_A. The inverse of $f \in Hom(G, A)$ for $\cdot$ is the homomorphism given by $g \mapsto f(g)^{-1}$ for all $g \in G$.

Dual group. Let

$$\Omega = t(\mathbb{C}^*) = \bigcup_{n=1}^{\infty} C_n. \tag{2.71}$$

Note that the homomorphism $\mathbb{Q} \to \mathbb{C}^*$ given by $q \mapsto e^{2\pi iq}$ has image Ω and kernel $\mathbb{Z}$; thus,

$$\Omega \cong \mathbb{Q}/\mathbb{Z},$$

It follows by problem 2.16 that every finitely generated subgroup of Ω is cyclic, and C_n is the unique subgroup of Ω of order n. For any finite abelian group A, the *dual group* of A is defined to be the abelian group

$$A^* \;=\; \mathit{Hom}(A, \Omega). \tag{2.72}$$

Note that if $\alpha\colon A \to C$ is any homomorphism of finite abelian groups, then there is an induced homomorphism $\alpha^*\colon C^* \to A^*$ given by

$$\alpha^*(f) \;=\; f \circ \alpha, \qquad \text{for any } \; f \text{ in } C^*.$$

Let $A^{**} = (A^*)^*$. There is a canonical homomorphism

$$\varepsilon_A\colon A \to A^{**} \quad \text{ given by } \quad \varepsilon_A(a)(f) = f(a), \tag{2.73}$$

for all $a \in A, f \in A^*$. We will see below that ε_A is an isomorphism for every finite abelian group A.

2.110. Prove that for every $n \in \mathbb{N}$,

$$C_n^* \;\cong\; C_n,$$

and the map $\varepsilon_{C_n}\colon C_n \to \mathbb{C}_n^{**}$ is an isomorphism.

2.111. Let G be a finite abelian group, and suppose there is $a \in G \setminus \{1_G\}$ such that a lies in every nontrivial subgroup of G. Prove that $|a| = p$ for some prime number p, and that G is cyclic of order a power of p. (If G is not assumed abelian, then G need not be cyclic, as the quaternion group Q_2 shows.)

2.112. For any finite abelian group A, prove that the map ε_A is injective. (Hint: If there is $a \in \mathit{ker}(\varepsilon_A)$ with $a \neq 1_A$, let B be a subgroup of A maximal such that $a \notin B$. Apply the preceding two problems to A/B to obtain a contradiction.)

2.113. Let A be any finite abelian group, and let B be any subgroup of A. Let $\pi\colon A \to A/B$ be the canonical projection given by $a \mapsto aB$, and let $\iota\colon B \to A$ be the inclusion map given by $b \mapsto b$ for $b \in B$. Let

$$B^\perp \;=\; \{\, f \in A^* \mid f(b) = 1 \text{ for all } b \in B \,\}, \tag{2.74}$$

which is a subgroup of A^*.

(i) Prove that π^* (mapping $(A/B)^*$ to A^*) is injective and that

$$\mathit{im}(\pi^*) \;=\; B^\perp = \mathit{ker}(\iota^*).$$

(Thus, $B^\perp \cong (A/B)^*$.)

(ii) Deduce that $\big|A^*\big| = \big|B^\perp\big| \cdot |\,im(\iota^*)| \leq \big|(A/B)^*\big| \cdot \big|B^*\big|$.

(iii) Deduce that $\big|A^*\big| \leq |A|$ by induction on $|A|$. It follows from this that $\big|A^{**}\big| \leq \big|A^*\big| \leq |A|$. Since $|A| \leq \big|A^{**}\big|$ by problem 2.112, it follows that

$$\big|A^{**}\big| \;=\; \big|A^*\big| \;=\; |A|,$$

and ε_A is an isomorphism $A \cong A^{**}$.

(iv) Prove that if B and C are subgroups of A with $A = B \times C$, then $A^* = C^\perp \times B^\perp$ Hence, by part (i),

$$A^* \;\cong\; (A/C)^* \times (A/B)^* \;\cong\; B^* \times C^*.$$

2.114. *Existence of cyclic decomposition.* Let A be any finite abelian group.

(i) Take any $f \in A^*$. Since $im(f)$ is a finite, hence cyclic, subgroup of Ω, there is $a \in A$ with $im(f) = \langle f(a)\rangle$. Prove that $|f|\,\big|\,|a|$.

(ii) Deduce from part (i) that $exp(A^*) \leq exp(A)$. It then follows that $exp(A^{**}) \leq exp(A^*)$. Since $exp(A^{**}) = exp(A)$ as $A^{**} \cong A$ by the preceding problem, it follows that

$$exp(A^*) \;=\; exp(A)$$

(iii) Take any $f \in A^*$ with $|f| = exp(A^*) = exp(A)$ and any $a \in A$ with $im(f) = \langle f(a)\rangle$. Prove that

$$A \;=\; \langle a\rangle \times ker(f)$$

and that $|a| = exp(A)$.

(iv) Deduce by induction on $|A|$ that A is a direct product of subgroups,

$$A \;=\; A_1 \times \ldots \times A_k,$$

with each A_i cyclic and $|A_{i+1}|\,\big|\,|A_i|$ for $i = 1, 2, \ldots, k-1$. This gives the existence part of the Fundamental Theorem. The uniqueness of the invariant factors will be covered in problem 2.116 below.

(v) Deduce that

$$A^* \cong A.$$

(However, there is no distinguished isomorphism between A and A^*, in contrast to the canonical isomorphism ε_A between A and A^{**}.)

2.115. *Duality.* Let A be any finite abelian group. Prove the following *duality theorem* for A and A^*: The map $B \mapsto B^\perp$ gives a one-to-one inclusion-reversing correspondence between the subgroups of A and the subgroups of A^*; moreover,

$$A^*/B^\perp \cong B^* \cong B \qquad \text{and} \qquad A/B \cong (A/B)^* \cong B^\perp.$$

(Inclusion-reversing means that if $B \subseteq C$, then $C^\perp \subseteq B^\perp$.) For a subgroup D of A^*, the corresponding subgroup of A is $\bigcap_{f\in D} ker(f)$.

For an abelian group A and any $n \in \mathbb{N}$, the *n-torsion subgroup of A* is

$$ {}_nA = \{\, a \in A \mid a^n = 1 \,\}. \tag{2.75}$$

It is a subgroup of A since A is abelian.

2.116. *Uniqueness of the invariant factors.* Let p be a prime number, and let $A = C_{p^{r_1}} \times \ldots \times C_{p^{r_k}}$, for any $r_i \in \mathbb{N}$. We may assume that $r_1 \geq r_2 \geq \ldots \geq r_i \geq \ldots \geq r_k$, so the p^{r_i} are invariant factors of A as in the Fundamental Theorem. Define nonnegative integers $s_0, s_1, s_2 \ldots, s_{r_1}$ by: $s_0 = 0$, and for $i \geq 1$, $p^{s_i} = \big|{}_{p^i}A\big|$. Prove that

$$s_1 = k = \big|\{\, j \mid r_j \geq 1 \,\}\big|,\ s_2 - s_1 = \big|\{\, j \mid r_j \geq 2 \,\}\big|,\ \ldots,$$
$$s_i - s_{i-1} = \big|\{\, j \mid r_j \geq i \,\}\big|,\ \ldots .$$

Hence,

$$\big|\{\, j \mid r_j = i \,\}\big| = 2s_i - s_{i-1} - s_{i+1} \quad \text{for } i = 1, 2, \ldots.$$

It follows that the invariant factors $p^{r_1}, \ldots, p^{r_k}$ of A are intrinsically (hence uniquely) determined by A, independent of the choice of cyclic direct product decomposition of A. This applies whenever A is p-primary, i.e, $A = A_{(p)}$. The uniqueness of the invariant factors for an arbitrary finite abelian group follows easily by using the primary decomposition in (2.65) and the Chinese Remainder Theorem as in (2.66).

2.117. Let A be a finite abelian group.

(i) Let B be a subgroup of A such that A/B is cyclic, say $A/B = \langle aB \rangle$, and suppose that $|A/B| = exp(A)$. Prove that $A = \langle a \rangle \times B$.

(ii) Now prove that if $c \in A$ with $|c| = exp(A)$, then there is a subgroup D of A with $A = \langle c \rangle \times D$. (Hint: Use duality and part (i).)

2.118. Let A and B be finite abelian groups. Let $n_1, \ldots, n_k$ be the invariant factors of A, and $m_1, \ldots, m_\ell$ the invariant factors of B. Prove that the following conditions are equivalent:

(a) B is isomorphic to a subgroup of A.

(b) $B \cong A/C$ for some subgroup C of A.

(c) $\ell \leq k$ and $m_i | n_i$ for $i = 1, 2, \ldots, \ell$.

Chapter 3

Rings

This chapter has problems on rings, starting with the basics and continuing to factorization theory in integral domains.

3.1. Rings, subrings, and ideals

Rings. A *ring* R is a nonempty set with two binary operations, addition, denoted $+$, and multiplication, denoted $\cdot$, such that

(i) $(R, +)$ is an abelian group;

(ii) $(r \cdot s) \cdot t = r \cdot (s \cdot t)$, for all $r, s, t \in R$;

(iii) $r \cdot (s+t) = (r \cdot s) + (r \cdot t)$ and $(r+s) \cdot t = (r \cdot t) + (s \cdot t)$, for all $r, s, t \in R$;

(iv) there is a (unique) multiplicative identity element $1_R \in R$ such that $1_R \cdot r = r \cdot 1_R = r$ for every $r \in R$.

The identity element for $+$ is denoted 0_R; so,

$$0_R + r = r + 0_R = r, \qquad \text{for every } r \in R.$$

A short calculation shows that

$$0_R \cdot r = r \cdot 0_R = 0_R, \qquad \text{for every } r \in R.$$

The additive inverse of r is denoted $-r$; thus,

$$r + -r = -r + r = 0_R.$$

There is a further operation of *subtraction* in R, defined by

$$r - s \;=\; r + (-s). \tag{3.1}$$

The ring R is *trivial* if $R = \{0_R\}$ or, equivalently, if $1_R = 0_R$. Ring R is *commutative* if $r \cdot s = s \cdot r$ for all $r, s \in R$. An element r of R is called a *unit of* R if there is $s \in R$ with $r \cdot s = s \cdot r = 1_R$. When such an s exists, it is unique, and is denoted r^{-1}. Let $R^* = \{\text{units of } R\}$, which is called the *group of units* of R. It is a group with the operation of multiplication in R, and the identity element of R^* is 1_R.

Not all authors require that a ring R have a multiplicative identity 1_R. However, all the rings considered in this book come naturally equipped with a multiplicative identity; so it is convenient to assume the presence of 1_R throughout. This is further justified since any "ring" without 1 (satisfying axioms (i)–(iii) above but not (iv)) is an ideal in a ring with 1. See problem 3.2 below.

If R and T are rings, a function $f\colon R \to T$ is a *ring homomorphism* if for all $r, s \in R$,

$$f(r+s) \;=\; f(r) + f(s) \quad \text{and} \quad f(r \cdot s) \;=\; f(r) \cdot f(s)$$
$$\textit{and} \quad f(1_R) \;=\; 1_T.$$

If in addition f is bijective, then f is a *ring isomorphism*. We write

$$R \cong T$$

when there is a ring isomorphism from R to T.

Here are some basic examples of rings:

- $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are commutative rings.
- $\mathbb{Z}_n$ is a commutative ring for any $n \in \mathbb{N}$, with the operations

  $$[i]_n + [j]_n \;=\; [i+j]_n \quad \text{and} \quad [i]_n \cdot [j]_n \;=\; [ij]_n,$$

  for all $i, j \in \mathbb{Z}$. Recall from problem 1.5 that these operations are well-defined. Note that $0_{\mathbb{Z}_n} = [0]_n$ and $1_{\mathbb{Z}_n} = [1]_n$.
- Let R be any ring. For any $n \in \mathbb{N}$, the set $M_n(R)$ of $n \times n$ matrices over R is a ring with respect to the usual operations of matrix addition and multiplication. Note that $0_{M_n(R)}$ is the matrix with all entries 0_R, and $1_{M_n(R)} = I_n$, the $n \times n$

identity matrix, with diagonal entries 1_R and all other entries 0_R. If $n \geq 2$, then $M_n(R)$ is noncommutative if R is nontrivial.

- Let R be any ring. Then the polynomial ring $R[X]$ and the formal power series ring $R[[X]]$ are again rings. (See pp. 77–79 below for the definitions.) $R[X]$ and $R[[X]]$ are commutative iff R is commutative.
- Let A be any abelian group, with the operation written additively. An *endomorphism* of A is a group homomorphism from A to A. Let

$$End(A) = \{\text{endomorphisms of } A\} = Hom(A, A). \quad (3.2)$$

For $f, g \in End(A)$, define $f + g\colon A \to A$ and $f \cdot g\colon A \to A$ by

$$(f + g)(a) = f(a) + g(a) \quad \text{and} \quad (f \cdot g)(a) = f(g(a))$$

for all $a \in A$. Since A is abelian, $f + g \in End(A)$. With these operations, $End(A)$ is a ring, called the *endomorphism ring of* A. The additive identity $0_{End(A)}$ is the trivial endomorphism sending every $a \in A$ to 0_A. The multiplicative identity $1_{End(A)}$ is the identity function id_A.

Ring notation: Let R be a ring. When the context is clear, we write 1 for 1_R and 0 for 0_R. Take any r, s and $r_1, r_2, \ldots, r_n$ in R. We write rs for $r \cdot s$; $r_1 + r_2 + \ldots + r_n$ for $(\ldots((r_1 + r_2) + r_3) + \ldots) + r_n$; and $r_1 r_2 \ldots r_n$ for $(\ldots((r_1 r_2) r_3)\ldots) r_n$. When parentheses are omitted, multiplication is done before addition or subtraction, e.g.,

$$rs - tu + vw \quad \text{means} \quad [(rs) - (tu)] + (vw).$$

For $r \in R$ and $n \in \mathbb{N}$, define

$$nr = \sum_{i=1}^{n} r \qquad \text{and} \qquad r^n = \prod_{i=1}^{n} r.$$

Also, set

$$0_{\mathbb{Z}} r = 0_R \qquad \text{and} \qquad r^{0_{\mathbb{Z}}} = 1_R.$$

Further, for $k \in \mathbb{Z}$ with $k < 0$, set $kr = (-k)(-r)$ and, if $r \in R^*$, $r^k = (r^{-1})^{-k}$.

Subrings. Let R be a ring. A nonempty subset T of R is a *subring* of R if $(T,+)$ is a subgroup of the additive group $(R,+)$, T is closed under multiplication, and $1_R \in T$ (so $1_T = 1_R$). Equivalently, T is a subring of R if for all $t, t' \in T$, we have $t - t' \in T$ and $tt' \in T$, and also $1_R \in T$. When we write "Let $T \subseteq R$ be rings," it is meant that T is a subring of R.

Ideals. Let R be a ring. A nonempty subset I of R is an *ideal* of R if $(I,+)$ is a subgroup of the additive group $(R,+)$ and for all $i \in I$ and $r \in R$ we have $ri \in I$ and $ir \in I$. The ideal I is a *proper ideal* if $I \neq R$; this holds iff $1_R \notin I$. An ideal I of R satisfies all the axioms for a ring except that (usually) it does not contain a multiplicative identity. If I is a proper ideal of R, then it is not a subring of R, since it does not contain 1_R. The *trivial ideal* of R is $\{0_R\}$. If I and J are ideals of R, then $I \cap J$ is an ideal of R, as is

$$I + J = \{i + j \mid i \in I,\, j \in J\}.$$

Also,

$$IJ = \Big\{ \sum_{k=1}^{n} i_k j_k \mid \text{each } i_k \in I,\, j_k \in J,\ n \in \mathbb{N} \Big\} \tag{3.3}$$

is an ideal of R. If S is any nonempty subset of R, then the *ideal of R generated by S* is

$$\Big\{ \sum_{i=1}^{n} r_i s_i t_i \mid \text{each } r_i, t_i, \in R,\ s_i \in S,\ n \in \mathbb{N} \Big\}.$$

This is the ideal of R containing S and lying in every other ideal of R containing S. It is also the intersection of all the ideals of R containing S.

Assume now that R is a commutative ring. For any $a_1, \ldots, a_n \in R$, the ideal of R generated by the a_i is

$$(a_1, a_2, \ldots, a_n) = \Big\{ \sum_{i=1}^{n} r_i a_i \mid \text{each } r_i \in R \Big\}. \tag{3.4}$$

An ideal generated by a single element is called a *principal ideal*. Thus, for $a \in R$, the principal ideal of R generated by a is

$$(a) = Ra = aR = \{ra \mid r \in R\}. \tag{3.5}$$

So, for $a, b \in R$,

$$(a)(b) = (ab) \subseteq (a) \cap (b) \qquad \text{and} \qquad (a) + (b) = (a, b).$$

We say that a *divides* b, denoted $a|b$, if there is $r \in R$ with $b = ra$. Note that $a|b$ iff $(b) \subseteq (a)$. We write $a \nmid b$ if a does not divide b.

Example 3.1. *Ideals of* $\mathbb{Z}$ *and* $\mathbb{Z}_n$. As noted in Example 2.1, every subgroup of the additive group $(\mathbb{Z}, +)$ has the form $k\mathbb{Z}$ for some unique $k \in \mathbb{Z}$ with $k \geq 0$. Note that $k\mathbb{Z}$ is the principal ideal (k) of $\mathbb{Z}$. Thus, for $\mathbb{Z}$, all additive subgroups are ideals, and all ideals are principal. This is also true for $\mathbb{Z}_n$ for any $n \in \mathbb{N}$, since (see Example 2.17) every additive subgroup of $\mathbb{Z}_n$ has the form $d\mathbb{Z}/n\mathbb{Z} = [d]_n\mathbb{Z}_n$ for some $d \in \mathbb{N}$ with $d|n$.

We give a careful definition of polynomials over a ring, since they are so essential in what follows. It is notationally convenient to define formal power series rings first.

Polynomial and formal power series rings. Let R be any ring. The *formal power series ring* $R[[X]]$ is defined to be the Cartesian product $\prod_{i=0}^{\infty} R$ with the operations defined as follows:
For $\alpha = (a_0, a_1, \dots, a_i, \dots)$ and $\beta = (b_0, b_1, \dots, b_i, \dots)$ in $R[[X]]$,

$$\begin{aligned} \alpha + \beta &= (a_0 + b_0,\ a_1 + b_1,\ \dots,\ a_i + b_i,\ \dots), \\ \alpha \cdot \beta &= (a_0 \cdot b_0,\ a_0 \cdot b_1 + a_1 \cdot b_0,\ \dots,\ \sum_{j=0}^{i} a_j \cdot b_{i-j},\ \dots). \end{aligned}$$

Straightforward calculations show that $R[[X]]$ is a ring, with

$$\begin{aligned} 0_{R[[X]]} &= (0, 0, \dots, 0, \dots), \quad 1_{R[[X]]} = (1_R, 0, 0, \dots, 0, \dots), \\ \text{and} \quad -(a_0, a_1, \dots, a_i, \dots) &= (-a_0, -a_1, \dots, -a_i, \dots). \end{aligned}$$

The customary notation is to write elements of this ring as formal infinite sums, writing $\sum_{i=0}^{\infty} a_i X^i$ for $(a_0, a_1, \dots, a_i, \dots)$. Then the operations are given as:

$$\begin{aligned} \sum_{i=0}^{\infty} a_i X^i + \sum_{i=0}^{\infty} b_i X^i &= \sum_{i=0}^{\infty} (a_i + b_i) X^i \\ \sum_{i=0}^{\infty} a_i X^i \cdot \sum_{i=0}^{\infty} b_i X^i &= \sum_{i=0}^{\infty} \Big(\sum_{j=0}^{i} a_j \cdot b_{i-j} \Big) X^i. \end{aligned}$$

This is suggestive notation for describing the ring operations. Note that there is no infinite summation of elements of R in these formulas,

so there are no convergence issues. The a_i are called the *coefficients* of $\sum_{i=0}^{\infty} a_i X^i$. Note that

$$\sum_{i=0}^{\infty} a_i X^i \;=\; \sum_{i=0}^{\infty} b_i X^i \quad \text{iff} \quad a_i = b_i \text{ for each } i.$$

The *polynomial ring* over R is the subring of $R[[X]]$ given by

$$R[X] \;=\; \Big\{ \sum_{i=0}^{\infty} a_i X^i \in R[[X]] \,\Big|\, \text{there is } k \in \mathbb{N} \text{ such that } a_i = 0 \text{ for all } i > k \Big\}. \tag{3.6}$$

When $a_i = 0$ for $i > k$, we write $\sum_{i=0}^{k} a_i X^i$ for $\sum_{i=0}^{\infty} a_i X^i$. A *constant polynomial* is one of the form

$$(a, 0, 0, 0, \ldots, 0, \ldots) \;=\; aX^0 + 0X + 0X^2 + \ldots + 0X^i + \ldots$$

for $a \in R$. The constant polynomials form a subring of $R[X]$, which is clearly isomorphic to R. We view R as a subring of $R[X]$ by identifying an element of R with the corresponding constant polynomial. Also, set

$$X \;=\; (0, 1, 0, 0, \ldots, 0, \ldots) \;=\; 0X^0 + 1X^1 + 0X^2 + \ldots + 0X^i + \ldots$$

in $R[X]$. Then,

$$X^0 \;=\; 1_{R[X]} = (1, 0, 0, \ldots, 0, \ldots), \quad X^1 \;=\; X \;=\; (0, 1, 0, \ldots, 0, \ldots)$$
$$X^2 \;=\; (0, 0, 1, 0, 0 \ldots, 0, \ldots), \quad X^3 \;=\; (0, 0, 0, 1, 0, 0, \ldots, 0, \ldots), \;\ldots,$$

and for $a \in R \subseteq R[X]$, we have

$$a \cdot X^i \;=\; (0, 0, \ldots, 0, \underset{i}{a}, 0, 0 \ldots, 0, \ldots).$$

Thus, the polynomial $\sum_{i=0}^{k} a_i X^i$ in $R[X]$ equals the sum of products $a_0 \cdot X^0 + a_1 \cdot X + a_2 \cdot X^2 + \ldots + a_k \cdot X^k$, justifying the summation notation for polynomials. For nonzero f in $R[X]$, we can write $f = \sum_{i=0}^{k} a_i X^i$ with $a_k \neq 0$. Then, the *degree* of f is $deg(f) = k$. (For $0_{R[X]}$ the degree is undefined.) Also, a_k is called the *leading coefficient* of f; if its leading coefficient is 1, then f is said to be *monic*. Take $f = \sum_{i=0}^{k} a_i X^i$ with $a_k \neq 0$ and $g = \sum_{j=0}^{\ell} b_j X^j$ with $b_\ell \neq 0$ in $R[X]$. Then,

$$deg(f + g) \;\leq\; max\big(deg(f), deg(g)\big), \;\text{ if } g \neq -f.$$

Also,

$$\begin{cases} \text{if } a_k b_\ell \neq 0, \text{ then } deg(fg) = k+\ell = deg(f) + deg(g) \\ \qquad \text{and } a_k b_\ell \text{ is the leading coefficient of } fg; \\ \text{if } a_k b_\ell = 0, \text{ then either } fg = 0 \text{ or } deg(fg) < deg(f) + deg(g). \end{cases}$$

3.2. *Adjoining a* 1 *to "rings" without* 1. Let T be a set with operations $+$ and $\cdot$ satisfying all the axioms of a ring except the existence of a 1. We show how to enlarge T to obtain a ring with a 1. Let $R = \mathbb{Z} \times T$, and define operations $+$ and $\cdot$ on R by

$$(k,t)+(\ell,s) = (k+\ell,\, t+s) \text{ and } (k,t)\cdot(\ell,s) = (k\ell,\, ks+\ell t+t\cdot s)$$

(i) Prove that R is a ring, with $1_R = (1_\mathbb{Z}, 0_T)$.

(ii) Let $T' = \{(0_\mathbb{Z}, t) \mid t \in T\} \subseteq R$. Prove that T' is an ideal of R, and that $T' \cong T$ (as rings without 1).

By embedding "rings" without 1 into rings (with 1) in this manner one can use results on rings to obtain analogues for "rings" without 1 for the infrequent occasions when they arise.

3.3. Let R be a ring. If I is an ideal of R, let $M_n(I)$ be the subset of the matrix ring $M_n(R)$ consisting of matrices all of whose entries are in I. It is easy to check that $M_n(I)$ is an ideal of $M_n(R)$. Now prove that every ideal of $M_n(R)$ has the form $M_n(I)$ for some ideal I of R.

3.4. Let R be a ring. Suppose there are $a, b \in R$ with $a \cdot b = 1$ but $b \cdot a \neq 1$. Prove that there are infinitely many elements $c \in R$ with $a \cdot c = 1$. (Equivalently, there are infinitely many $d \in R$ with $a \cdot d = 0$.)

3.5. Let R be a ring, and suppose that $a^3 = a$ for every a in R. Prove that R is commutative.

3.6. *Idempotents.* An element e of a ring R is said to be *idempotent* if $e^2 = e$. The idempotent e is *nontrivial* if $e \neq 0$ and $e \neq 1$. When e is idempotent, $1 - e$ is also idempotent. It is easy to check that eRe $(= \{ere \mid r \in R\})$ is a ring under the operations inherited from R, with $1_{eRe} = e$. But eRe is not considered a subring of R when $e \neq 1$ since its multiplicative identity is not 1_R. Note that if I is an ideal of R, then eIe $(= \{eae \mid a \in I\})$ is an ideal of eRe, and $eIe = I \cap eRe$

Let

$$ReR = \Big\{ \sum_{i=1}^{n} r_i e s_i \mid r_i,\, s_i \in R,\, n \in \mathbb{N} \Big\},$$

which is the ideal of R generated by e. Suppose that $ReR = R$. Prove that if I is an ideal of R, then I is generated, as an ideal of R, by eIe. Conversely, if J is an ideal of eRe, let I be the ideal of R generated by J. Prove that $I \cap eRe = J$. Thus, when $ReR = R$, there is a natural inclusion-preserving one-to-one correspondence between the ideals of eRe and the ideals of R.

Direct products. Let $\{R_i \mid i \in I\}$ be a collection of rings. The *direct product* of the R_i is the Cartesian product $\prod_{i\in I} R_i$, made into a ring with componentwise operations. For each $j \in I$, the projection map $\pi_j \colon \prod_{i\in I} R_i \to R_j$, taking an element to its j-th component, is a ring homomorphism. The universal mapping property for the direct product is: For any ring T and any family of ring homomorphisms $\alpha_i \colon T \to R_i$ for all $i \in I$, there is a unique ring homomorphism $\beta \colon T \to \prod_{i\in I} R_i$ such that $\alpha_i = \pi_i \circ \beta$ for every i.

3.7. Let R_1 and R_2 be rings, and let $R_1 \times R_2$ be their direct product.

(i) It is easy to see that if I_i is an ideal of R_i for $i = 1, 2$, then $I_1 \times I_2$ is an ideal of the direct product $R_1 \times R_2$. Prove that every ideal of $R_1 \times R_2$ has the form $I_1 \times I_2$ for some ideals I_i of R_i.

(ii) Prove that $(R_1 \times R_2)^* \cong R_1^* \times R_2^*$, a direct product of groups.

The analogues to parts (i) and (ii) hold for arbitrary direct products $\prod_{i\in I} R_i$ of rings. Note that, with the Chinese Remainder Theorem for rings (see problem 3.17 below), part (ii) yields the group isomorphism given earlier in (2.16): For any $m, n \in \mathbb{N}$ with $gcd(m, n) = 1$,

$$\mathbb{Z}_{mn}^* \;\cong\; (\mathbb{Z}_m \times \mathbb{Z}_n)^* \;\cong\; \mathbb{Z}_m^* \times \mathbb{Z}_n^*.$$

The *center* of a ring R is

$$Z(R) \;=\; \{a \in R \mid ar = ra \text{ for every } r \in R\}, \tag{3.7}$$

which is a commutative subring of R. For example, for any $n \in \mathbb{N}$, the center of $M_n(R)$ consists of all diagonal matrices with the same element of $Z(R)$ for each diagonal entry. Thus, $Z(M_n(R)) \cong Z(R)$.

3.8. Let R be a ring, and let e be an idempotent of R with $e \in Z(R)$. Let $f = 1 - e$, which is also a central idempotent of R. Thus, $Re = eR = eRe$ and $Rf = fR = fRf$. We have seen in problem 3.6 that eRe and fRf are rings. Prove that

$$R \cong eRe \times fRf.$$

Note, conversely, that if $R = R_1 \times R_2$ for some rings R_1 and R_2, then for $e = (1_{R_1}, 0_{R_2})$ and $f = 1_R - e$, we have $e \in Z(R)$ and $R_1 \cong eRe$ and $R_2 \cong fRf$.

A *field* F is a commutative ring with $1_F \neq 0_F$ such that

$$F^* = F \setminus \{0_F\}.$$

Equivalently, F is a field if it is a nontrivial commutative ring in which the only ideals are F and $\{0_F\}$. For example, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, but $\mathbb{Z}$ is not a field. Also, for $n \in \mathbb{N}$, the ring $\mathbb{Z}_n$ is a field iff n is a prime number. A subring of field F that is also a field is called a *subfield* of F.

3.9. This problem gives a way of constructing $\mathbb{C}$ from $\mathbb{R}$. Let

$$\mathcal{C} = \left\{ \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \;\middle|\; a, b \in \mathbb{R} \right\} \subseteq M_2(\mathbb{R}).$$

(i) Prove that $\mathcal{C}$ is a subring of $M_2(\mathbb{R})$.

(ii) Prove that $\mathcal{C}$ is a field.

(iii) Let $\mathcal{R} = \left\{ \left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right) \;\middle|\; r \in \mathbb{R} \right\} \subseteq \mathcal{C}$, and let $\mathbf{i} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \in \mathcal{C}$. It is easy to check that $\mathcal{R}$ is a subring of $\mathcal{C}$ and that $\mathcal{R} \cong \mathbb{R}$. Prove further that every element of $\mathcal{C}$ is uniquely expressible as $\alpha + \beta\mathbf{i}$ with $\alpha, \beta \in \mathcal{R}$, and that $\mathbf{i}^2 = -1_{\mathcal{R}}$. Thus, we can identify $\mathcal{C}$ with the field of $\mathbb{C}$ of complex numbers.

Ring automorphisms. Let R be a ring. A (*ring*) *automorphism* of R is a ring isomorphism from R onto R. Let

$$Aut(R) = \{\text{ring automorphisms of } R\}. \tag{3.8}$$

Clearly, $Aut(R)$ is a subgroup of $\Sigma(R)$. For any $u \in R^*$, the map *conjugation by* u,

$$\gamma_u \colon R \to R \qquad \text{given by} \qquad r \mapsto uru^{-1} \tag{3.9}$$

is easily seen to be a ring automorphism of R. (Its inverse map is $\gamma_{u^{-1}}$.) Such a conjugation map is called an *inner automorphism* of R. Note that for $u, v \in R^*$, we have $\gamma_u \circ \gamma_v = \gamma_{uv}$. Thus, there is a group homomorphism

$$\psi\colon R^* \to Aut(R) \qquad \text{given by} \qquad u \mapsto \gamma_u.$$

Clearly, $ker(\psi) = Z(R) \cap R^*$. But ψ is often not surjective. For example, complex conjugation on $\mathbb{C}$ is a ring automorphism that is not inner.

3.10. Let $T = \left\{ \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right) \mid a, b, c \in \mathbb{Q} \right\}$, the ring of upper triangular matrices in $M_2(\mathbb{Q})$.

(i) Determine all the idempotent elements of T.

(ii) Determine all the inner automorphisms of T.

(iii) Determine whether every ring automorphism of T is inner.

3.11. *Quaternions.* A *division ring* is a ring R such that $1_R \neq 0_R$ and every nonzero element of R is a unit. Commutative division rings are fields, but noncommutative division rings are more difficult to find. This problem gives a construction of Hamilton's quaternions, $\mathbb{H}$, which were the first discovered example of a division ring that is not a field. We work in $M_2(\mathbb{C})$. For $\alpha = a + bi \in \mathbb{C}$, where $a, b \in \mathbb{R}$ and $i^2 = -1$, let $\overline{\alpha} = a - bi$, the complex conjugate of α. Recall that

$$\alpha\,\overline{\alpha} = a^2 + b^2 = |\alpha|^2 \in \mathbb{R};$$

thus, $\alpha\overline{\alpha} \geq 0$ with equality holding iff $\alpha = 0$. Let

$$\mathbb{H} = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \mid \alpha, \beta \in \mathbb{C} \right\} \subseteq M_2(\mathbb{C}). \tag{3.10}$$

Note that $\mathbb{H}$ contains the following four elements:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

which satisfy the following equations:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}, \tag{3.11}$$

$$\mathbf{ij} = \mathbf{k} = -\mathbf{ji}, \quad \mathbf{jk} = \mathbf{i} = -\mathbf{kj}, \quad \mathbf{ki} = \mathbf{j} = -\mathbf{ik} \tag{3.12}$$

(i) Prove that $\mathbb{H}$ is a subring of $M_2(\mathbb{C})$, and that $1_{\mathbb{H}} = \mathbf{1}$.

(ii) For $c \in \mathbb{C}$ and $A \in M_2(\mathbb{C})$, let cA denote the matrix obtained by multiplying all the entries by c. Note that for $r, s, t, u \in \mathbb{R}$,

$$r\mathbf{1} + s\mathbf{i} + t\mathbf{j} + u\mathbf{k} = \begin{pmatrix} r+si & t+ui \\ -t+ui & r-si \end{pmatrix}.$$

Deduce that every $h \in \mathbb{H}$ is uniquely expressible as

$$h = r\mathbf{1} + s\mathbf{i} + t\mathbf{j} + u\mathbf{k}$$

with $r, s, t, u \in \mathbb{R}$. Thus, $\mathbb{H}$ is a 4-dimensional vector space over $\mathbb{R}$ with base $\{\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$.

(iii) Equations (3.11) show that $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathbb{H}^*$ with

$$\mathbf{i}^{-1} = -\mathbf{i}, \quad \mathbf{j}^{-1} = -\mathbf{j}, \quad \text{and} \quad \mathbf{k}^{-1} = -\mathbf{k}.$$

Note that for all $\alpha, \beta \in \mathbb{C}$,

$$\mathbf{i}\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}\mathbf{i}^{-1} = \begin{pmatrix} \alpha & -\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix}$$

and

$$\mathbf{j}\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}\mathbf{j}^{-1} = \begin{pmatrix} \overline{\alpha} & \overline{\beta} \\ -\beta & \alpha \end{pmatrix}.$$

Deduce that $Z(\mathbb{H}) = \{\, r\mathbf{1} \mid r \in \mathbb{R}\} \cong \mathbb{R}$.

(iv) Define a function $\sigma\colon \mathbb{H} \to \mathbb{H}$ by

$$\sigma\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} = \begin{pmatrix} \overline{\alpha} & -\beta \\ \overline{\beta} & \alpha \end{pmatrix}.$$

Thus, for all $r, s, t, u \in \mathbb{R}$,

$$\sigma(r\mathbf{1} + s\mathbf{i} + t\mathbf{j} + u\mathbf{k}) = r\mathbf{1} - s\mathbf{i} - t\mathbf{j} - u\mathbf{k}.$$

Prove that for all $h, \ell \in \mathbb{H}$

$$\sigma(h+\ell) = \sigma(h) + \sigma(\ell) \quad \text{and} \quad \sigma(h\ell) = \sigma(\ell)\sigma(h)$$

(note the reversal of order of terms in the product), and

$$h\sigma(h) = \sigma(h)h = det(h)\mathbf{1}, \tag{3.13}$$

where $det(h)$ is the determinant of the 2×2 matrix h. The map σ is called an *involution* on $\mathbb{H}$, since it is an *anti-automorphism* (i.e., a bijection that preserves addition but reverses the order of multiplication) with $\sigma \circ \sigma = id_{\mathbb{H}}$.

(v) Note that for $h = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \in \mathbb{H}$, we have

$$det(h) \;=\; \alpha\overline{\alpha} + \beta\overline{\beta} \;=\; |\alpha|^2 + |\beta|^2,$$

which lies in $\mathbb{R}$ and is nonnegative. Also, for $r, s, t, u \in \mathbb{R}$,

$$det(r\mathbf{1} + s\mathbf{i} + t\mathbf{j} + u\mathbf{k}) \;=\; r^2 + s^2 + t^2 + u^2.$$

Deduce that any nonzero $h \in H$ has a multiplicative inverse given by

$$h^{-1} \;=\; det(h)^{-1}\sigma(h). \tag{3.14}$$

Thus, $\mathbb{H}^* = \mathbb{H} \setminus \{0\}$, showing that $\mathbb{H}$ is a division ring.

Note: Let $Q = \{\pm\mathbf{1}, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\} \subseteq \mathbb{H}^*$. The formulas in (3.11) and (3.12) show that Q is a subgroup of $\mathbb{H}^*$. This Q is a nonabelian group of order 8 called the *quaternion group*. It is isomorphic to the group Q_2 of problem 2.9.

3.12. This problem describes some of the remarkable geometric properties embedded in the quaternions. Keep the notation of the preceding problem. Let

$$P \;=\; \{s\mathbf{i} + t\mathbf{j} + u\mathbf{k} \mid s, t, u \in \mathbb{R}\} \;\subseteq\; \mathbb{H}.$$

This P is sometimes called the "pure part" of $\mathbb{H}$, in analogy with the purely imaginary part of $\mathbb{C}$. Note that every $h \in \mathbb{H}$ is uniquely expressible as

$$h \;=\; r\mathbf{1} + p \quad \text{with } r \in \mathbb{R} \text{ and } p \in P;$$

$r\mathbf{1}$ is then called the "real part" of h, and p is called the "pure part" of h. Note that

$$r\mathbf{1} \;=\; \tfrac{1}{2}\big(h + \sigma(h)\big) \quad \text{and} \quad p \;=\; \tfrac{1}{2}\big(h - \sigma(h)\big).$$

The set P is the 3-dimensional subspace of the 4-dimensional real vector space $\mathbb{H}$ spanned by $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. But P is not closed under multiplication, so it cannot be a subring nor an ideal of $\mathbb{H}$. Note also that for $p = s\mathbf{i} + t\mathbf{j} + u\mathbf{k} \in P$, since $\sigma(p) = -p$, we have

$$p^2 \;=\; -p\sigma(p) \;=\; -\,det(p)\mathbf{1} \;=\; -(s^2 + t^2 + u^2)\mathbf{1}. \tag{3.15}$$

There is significant geometric information encoded in the multiplication of elements of P. We can identify P with the 3-dimensional column vector space $\mathbb{R}^3$ over $\mathbb{R}$ by the correspondence

$$p = p_1\mathbf{i} + p_2\mathbf{j} + p_3\mathbf{k} \longleftrightarrow \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

(which is an $\mathbb{R}$-vector space isomorphism). On $\mathbb{R}^3$ we have the operations the *dot product* $\bullet\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ given by

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \bullet \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = p_1q_1 + p_2q_2 + p_3q_3, \tag{3.16}$$

and the *cross product* $\times\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \times \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} p_2q_3 - p_3q_2 \\ p_3q_1 - p_1q_3 \\ p_1q_2 - p_2q_1 \end{pmatrix}, \tag{3.17}$$

and also the Euclidean norm $\|\ \|\colon \mathbb{R}^3 \to \mathbb{R}$ given by

$$\left\| \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \right\| = \sqrt{p_1^2 + p_2^2 + p_3^2}\,. \tag{3.18}$$

(i) Prove that for all $p, q \in P$,

$$pq = -(p \bullet q)\mathbf{1} + p \times q. \tag{3.19}$$

That is, the real part of pq is -1 times the dot product of the corresponding vectors, and its pure part is the element of P corresponding to the cross product of the corresponding vectors. In particular, note that

$$p^2 = -(p \bullet p)\mathbf{1} = -\|p\|^2\mathbf{1}. \tag{3.20}$$

(ii) Take any $p, q \in P$. Prove that

$$qp = -pq \quad \text{iff} \quad p \perp q \text{ (i.e., } p \bullet q = 0).$$

(Recall that we always have $p \bullet q = q \bullet p$ and $p \times q = -q \times p$.) Prove also that

$$qp = pq \quad \text{iff} \quad p \text{ and } q \text{ are parallel}$$

(i.e., $p = sq$ for some $s \in \mathbb{R}$ or $q = 0$).

(iii) For $h \in \mathbb{H}$, prove that if $h^2 \in \mathbb{R}\mathbf{1}$ then $h \in \mathbb{R}\mathbf{1}$ or $h \in P$. (To compute h^2 you can write $h = r\mathbf{1} + p$ with $r \in \mathbb{R}$ and $p \in P$.) Prove that if $h^2 = s\mathbf{1}$ with $s \leq 0$ in $\mathbb{R}$, then $h \in P$.

(iv) Let $h \in \mathbb{H}^*$, and let $\gamma_h \colon \mathbb{H} \to \mathbb{H}$ be the conjugation by h inner automorphism given by $\ell \mapsto h\ell h^{-1}$, as in (3.9). In view of the formula (3.14) for h^{-1}, we have

$$\gamma_h(\ell) \;=\; \tfrac{1}{det(h)} h\ell\sigma(h)$$

for any $\ell \in \mathbb{H}$. Prove that $\gamma_h \circ \sigma = \sigma \circ \gamma_h$, and that γ_h maps P to P and $\mathbb{R}\mathbf{1}$ to $\mathbb{R}\mathbf{1}$.

(v) Let $h = r\mathbf{1} + p \in \mathbb{H}^*$ with $r \in \mathbb{R}$ and $p \in P$. It follows from part (iv) and (3.19) that the action of the automorphism γ_h on P preserves dot products (and cross products). Hence, it must be a rigid motion of P (cf. problem 2.11), sending the origin to itself. But there is a much more explicit description of γ_h; it is a rotation: Specifically, assume that $p \neq 0$ (since otherwise $\gamma_h = id_{\mathbb{H}}$). Then, γ_h gives a rotation of P about the p-axis, through an angle 2θ where $\theta > 0$ and

$$cos(\theta) \;=\; r\big/\sqrt{det(h)} \quad \text{and} \quad sin(\theta) \;=\; \sqrt{-p^2}\big/\sqrt{det(h)}\,.$$

The p-axis is the line $\{tp \mid t \in \mathbb{R}\}$ through the point p and the origin. The rotation is counterclockwise through 2θ viewed from p toward the origin. Such a θ exists since

$$\big(r\big/\sqrt{det(h)}\,\big)^2 + \big(\sqrt{-p^2}\big/\sqrt{det(h)}\,\big)^2 \\ = (r^2 - p^2)/\,det(h) \;=\; 1.$$

Now prove this description of the action of γ_h in the particular case where $h = c\mathbf{1} + s\mathbf{i}$, with $c, s \in \mathbb{R}$ such that $c^2 + s^2 = 1$, so $cos(\theta) = c$ and $sin(\theta) = s$. (Here is one way to understand this: Let $C = \{r\mathbf{1} + t\mathbf{i} \mid r, t \in \mathbb{R}\}$, which is a subring of $\mathbb{H}$ isomorphic to $\mathbb{C}$; then, $\mathbb{H} = C + C\mathbf{j}$. Left multiplication by $h = c\mathbf{1} + s\mathbf{i}$ does not send P to itself, but acts on C and on $C\mathbf{j}$ by rotation by θ. Right multiplication by $\sigma(h) = c\mathbf{1} - s\mathbf{i}$ acts on C by rotation by $-\theta$, but on $C\mathbf{j}$ by rotation by θ, as $\mathbf{j}\sigma(h) = h\mathbf{j}$. The composition of these two operations is the identity on the $\mathbf{i}$-axis and rotation by 2θ on $C\mathbf{j}$.)

(vi) Take any $u \in P$ with $det(u) = 1$. Then choose any $v \in P$ with $v \perp u$ (i.e., $v \bullet u = 0$) and $det(v) = 1$. Let $\mathbf{i}' = u$, $\mathbf{j}' = v$, and $\mathbf{k}' = uv = u \times v$ (see (3.19)). Prove that $\mathbf{i}'$, $\mathbf{j}'$, and $\mathbf{k}'$

satisfy the same identities as $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ in (3.11) and (3.12). Prove also that $\mathbf{i}'$, $\mathbf{j}'$, and $\mathbf{k}'$ are $\mathbb{R}$-linearly independent in P, so form an $\mathbb{R}$-base of P.

(vii) Deduce from parts (v) and (vi) that for any $h = r\mathbf{1} + p \in \mathbb{H}^*$ with $r \in \mathbb{R}$ and $p \in P \backslash \{0\}$ the action of γ_h on P is a rotation, as described in part (v).

Note: Because of the result in part (vii), quaternions are used in computer graphics. Any orientation-preserving rigid motion of $\mathbb{R}^3$ fixing the origin is known to be a rotation about a line. (See problem 4.105(v) below.) Any rotation about a line through the origin in $\mathbb{R}^3$ corresponds to the action of γ_h on P for some $h \in \mathbb{H}^*$. A second rotation about a (possibly different) line through the origin corresponds to some γ_ℓ. The composition of these rotations is again a rotation, typically about some new line through the origin. This is not immediately obvious geometrically, but follows at once from the quaternionic description: The composition of the rotations corresponds to $\gamma_\ell \circ \gamma_h = \gamma_{\ell h}$; moreover, we can read off from $h\ell$ the axis and angle of rotation for the composition. For more on the quaternions, see the excellent article by M. Koecher and R. Remmert, Chapter 7 in Ebbinghaus et al. [**6**].

3.13. Problem 3.12 parts (v) and (vii) show that the composition of two rotations of $\mathbb{R}^3$ about axes through the origin is again a rotation, typically about some different axis, and they show how one can use the product of elements in $\mathbb{H}$ to calculate the composite rotation. This problem gives a geometric approach to the same result. The key idea is to express rotations as compositions of reflections. In $\mathbb{R}^3 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}$, let $\mathcal{S}$ be the unit sphere,

$$\mathcal{S} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x^2 + y^2 + z^2 = 1 \right\},$$

and let $\mathrm{O} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, the origin. A *great circle* on $\mathcal{S}$ is a circle with center O (necessarily of radius 1). Two distinct points on $\mathcal{S}$ are *antipodal* if the line they determine passes through O. Take two points A and B on $\mathcal{S}$ that are not antipodal. The plane $\mathcal{P}_{AB}$ determined by A, B, and O intersects $\mathcal{S}$ in the unique great circle $\mathcal{C}_{AB}$ on $\mathcal{S}$ containing both A and B. The *spherical arc* $\mathcal{A}_{AB}$ connecting A and B on $\mathcal{S}$ is

the shorter arc on $\mathcal{C}_{AB}$ between A and B. (The arc $\mathcal{A}_{AB}$ is known to be the shortest path between A and B along $\mathcal{S}$.)

(i) Let A and B be two distinct nonantipodal points on $\mathcal{S}$, and let $\rho_{AB}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection map across the plane $\mathcal{P}_{AB}$. Clearly, ρ_{AB} maps $\mathcal{S}$ bijectively to itself. Let C be a point of $\mathcal{S}$ not lying on $\mathcal{C}_{AB}$. Then, A, B, and C determine a "spherical triangle" whose sides are the spherical arcs $\mathcal{A}_{AB}$, $\mathcal{A}_{BC}$, and $\mathcal{A}_{CA}$. The angle $\angle ABC$ of this triangle at B is defined to be the angle between the planes $\mathcal{P}_{AB}$ and $\mathcal{P}_{BC}$, or, equivalently, the angle between the tangent lines to $\mathcal{A}_{AB}$ and to $\mathcal{A}_{BC}$ at B. Now, let

$$\tau_{ABC} \;=\; \rho_{BC} \circ \rho_{AB}\colon \mathbb{R}^3 \to \mathbb{R}^3.$$

Prove that τ_{ABC} is a rotation about the axis determined by B and O. Prove further that the angle of rotation of τ_{ABC}, moving from $\mathcal{P}_{AB}$ in the direction toward $\mathcal{P}_{BC}$, is twice the angle $\angle ABC$.

(ii) Let A, B, and C be the vertices of a spherical triangle on $\mathcal{S}$, as in part (i). Prove that

$$\tau_{CAB} \circ \tau_{BCA} \circ \tau_{ABC} \;=\; id_{\mathbb{R}^3}\,.$$

Thus,

$$\tau_{BCA} \circ \tau_{ABC} \;=\; {\tau_{CAB}}^{-1},$$

which expresses the composition of rotations τ_{BCA} and τ_{ABC} as a rotation.

The preceding problem suggests a process for realizing the composition of two rotations through the origin in $\mathbb{R}^3$: Given rotations τ_1 and τ_2 about different axes through O, let B (resp. C) be a point of $\mathcal{S}$ on the axis of τ_1 (resp. τ_2). Let $\mathcal{P}_i$ be the plane obtained by rotating plane $\mathcal{P}_{BC}$ through half the angle of rotation of τ_i, for $i = 1, 2$. Then, $\mathcal{P}_1 \cap \mathcal{P}_2$ is the axis of rotation of $\tau_2 \circ \tau_1$; moreover, if $\mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{S} = \{A, A'\}$, then the angle of rotation of $\tau_2 \circ \tau_1$ is twice $\angle CAB$ or $\angle CA'B$.

3.14. *Automorphisms of the quaternions.* Let $\alpha\colon \mathbb{H} \to \mathbb{H}$ be any ring automorphism of the quaternions, such that $\alpha(r\mathbf{1}) = r\mathbf{1}$ for all $r \in \mathbb{R}$. Prove that there is an $h \in \mathbb{H}^*$ such that α is the inner automorphism

conjugation by h. (The assumption that $\alpha(r\mathbf{1}) = r\mathbf{1}$ actually holds for all automorphisms of $\mathbb{H}$ by problem 5.65 below.)

3.2. Factor rings and ring homomorphisms

Factor rings. Let R be a ring, and let I be an ideal of R. The *factor ring of R modulo I* is

$$R/I \;=\; \{r+I \mid r \in R\} \quad \text{where} \quad r+I = \{r+i \mid i \in I\}, \tag{3.21}$$

with operations given by

$$(r+I)+(s+I) \;=\; (r+s)+I \;\text{ and }\; (r+I)\cdot(s+I) \;=\; (r\cdot s)+I.$$

Thus, the elements of R/I are the cosets of the additive subgroup I in the additive group R, and the additive group of R/I is the factor group of the additive group R modulo its normal subgroup I. Note that

$$r+I \;=\; s+I \quad \text{iff} \quad r-s \in I.$$

It is easy to check that the operations on R/I are well-defined, and that with these operations, R/I is a ring, with $0_{R/I} = 0_R + I = I$, $1_{R/I} = 1_R + I$, and $-(r+I) = (-r)+I$, so $(r+I)-(s+I) = (r-s)+I$. The canonical projection $\pi\colon R \to R/I$ given by $r \mapsto r+I$ is a surjective ring homomorphism.

Recall that for any rings R and T, a *ring homomorphism* from R to T is a function $f\colon R \to T$ such that $f(r+s) = f(r)+f(s)$ and $f(r\cdot s) = f(r)\cdot f(s)$ for all $r, s \in R$, and $f(1_R) = 1_T$. Note that the *image* of f,

$$im(f) \;=\; \{f(r) \mid r \in R\}$$

is a subring of T and the *kernel* of f,

$$ker(f) \;=\; \{r \in R \mid f(r) = 0_T\} \tag{3.22}$$

is an ideal of R. Note also that the ring homomorphism f is in particular a group homomorphism from the addititive group $(R,+)$ to $(T,+)$, and all the results about group homomorphisms apply. For example, $f(0_R) = 0_T$ and $f(-r) = -f(r)$ for all $r \in R$.

The basic homomorphism and isomorphism theorems for groups each have analogues for rings, which we now recall. They are all easy to prove by first applying the corresponding group theorem to the underlying additive group homomorphism of a ring homomorphism.

The *Fundamental Homomorphism Theorem* (FHT) for rings says: Let R, T be rings, and let $f\colon R \to T$ be a ring homomorphism. Let I be an ideal of R, and let $\pi\colon R \to R/I$ be the canonical projection. Suppose that $I \subseteq ker(f)$. Then, there is a unique induced homomorphism $g\colon R/I \to T$ such that $f = g \circ \pi$. Moreover, $im(g) = im(f)$ and $ker(g) = ker(f)/I$. In particular (taking $I = ker(f)$), we have the ring isomorphism

$$R/ker(f) \cong im(f).$$

The *First Isomorphism Theorem* for rings says: Let I be an ideal of a ring T, and let R be any subgroup of T. Then, $R + I$ $(= \{r + i \mid r \in R,\, i \in I\})$ is a subring of T and $I \cap R$ is an ideal of R, and

$$R/(I \cap R) \;\cong\; (R + I)/I.$$

This ring isomorphism is given by $r + (I \cap R) \mapsto r + I$.

The *Second Isomorphism Theorem* for rings says: Let I and J be ideals of a ring R with $I \subseteq J$. Then, there is a well-defined surjective ring homomorphism $f\colon R/I \to R/J$ given by $r + I \mapsto r + J$ for $r \in R$. Also, $ker(f) = J/I = \{j + I \mid j \in J\}$, which is an ideal of R/I, and f induces a ring isomorphism

$$R/I\big/J/I \;\cong\; R/J,$$

given by $(r + I) + J/I \mapsto r + J$. Moreover, every ideal of R/I has the form J/I for some ideal J of R with $J \supseteq I$.

The *Correspondence Theorem* for rings says: Let $f\colon R \to R'$ be a surjective homomorphism of rings. Then, in the one-to-one correspondence between the set of subgroups of the additive group of R and the subgroups of the additive group of R' given by the Correspondence Theorem for groups, the subrings of R containing $ker(f)$ correspond to the subrings of R'; likewise, the ideals I of R containing $ker(f)$ correspond to the ideals I' of R'. When $I \leftrightarrow I'$, we have a ring isomorphism $R/I \cong R'/I'$.

Example 3.15. Let R be a ring, and let I be an ideal of R. For $r \in R$, let $\overline{r} = r + I$, which is the image of r in R/I.

(i) For any $n \in \mathbb{N}$ the map $M_n(R) \to M_n(R/I)$ given by $(r_{ij}) \mapsto (\overline{r_{ij}})$ is a surjective ring homomorphism with kernel $M_n(I) = \big\{(r_{ij}) \mid \text{each } r_{ij} \in I\big\}$. Hence, by the FHT,

$$M_n(R)\big/M_n(I) \;\cong\; M_n(R/I).$$

(ii) The map of polynomial rings $R[X] \to (R/I)[X]$ given by

$$\sum_{i=0}^{k} a_i X^i \;\mapsto\; \sum_{i=0}^{k} \overline{a_i} X^i$$

is a surjective ring homomorphism with kernel

$$IR[X] \;=\; \Big\{ \sum_{i=0}^{k} b_i X^i \mid \text{each } b_i \in I\Big\}.$$

Hence, by the FHT,

$$R[X]\big/IR[X] \;\cong\; (R/I)[X].$$

(iii) Let $R_1, R_2, \ldots R_n$ be rings. We know (see problem 3.7) that every ideal of $R_1 \times \ldots \times R_n$ has the form $I_1 \times \ldots \times I_n$, where each I_j is an ideal of R_j. The map

$$R_1 \times \ldots \times R_n \;\to\; (R_1/I_1) \times \ldots \times (R_n/I_n)$$

given by $(r_1, \ldots, r_n) \mapsto (r_1 + I_1, \ldots, r_n + I_n)$ is a surjective ring homomorphism with kernel $I_1 \times \ldots \times I_n$. Hence, by the FHT,

$$\big(R_1 \times \ldots \times R_n\big)\big/\big(I_1 \times \ldots \times I_n\big) \cong (R_1/I_1) \times \ldots \times (R_n/I_n).$$

Example 3.16. For any $n \in \mathbb{N}$, the surjective group homomorphism $\mathbb{Z} \to \mathbb{Z}_n$ given by $i \mapsto [i]_n$ is also a ring homomorphism. Hence, the group isomorphism given by the FHT for groups,

$$\mathbb{Z}_n \;\cong\; \mathbb{Z}/n\mathbb{Z}$$

is also a ring isomorphism, by the FHT for rings. Likewise, for any $n_1, \ldots, n_k \in \mathbb{N}$ with $gcd(n_i, n_j) = 1$ whenever $i \neq j$, the surjective group homomorphism $\mathbb{Z} \to \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_k}$ given by $i \mapsto ([i]_{n_1}, \ldots, [i]_{n_k})$ is actually a ring homomorphism, so the induced group isomorphism of the Chinese Remainder Theorem,

$$\mathbb{Z}_{n_1 \ldots n_k} \;\cong\; \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_k}, \tag{3.23}$$

(see Example 2.19) is also a ring isomorphism. The next problem gives a generalization of this theorem.

3.17. *Chinese Remainder Theorem for commutative rings.* Let R be a commutative ring.

(i) Let I and J be ideals of R such that $I + J = R$. Prove that $IJ = I \cap J$ and that there is a ring isomorphism

$$R/IJ \cong R/I \times R/J.$$

(Hint: We have $1 = i + j$ for some $i \in I$ and $j \in J$. For any $r, s \in R$, note that $jr + is$ maps to $r + I$ in R/I and to $s + J$ in R/J.)

(ii) Building on part (i), suppose that $I_1, \ldots, I_k$ are ideals of R such that $I_i + I_j = R$ whenever $i \neq j$. Prove by induction on k that

$$I_1 I_2 \ldots I_k = I_1 \cap I_2 \cap \ldots \cap I_k$$

and that there is a ring isomorphism

$$R/I_1 I_2 \ldots I_k \cong R/I_1 \times R/I_2 \times \ldots \times R/I_k. \tag{3.24}$$

This is the *Chinese Remainder Theorem for commutative rings.* Note that the Chinese Remainder Theorem for the integers is a special case.

An element r of a ring R is said to be *nilpotent* if $r^n = 0$ for some $n \in \mathbb{N}$.

3.18. Let R be a commutative ring, and let

$$\mathcal{N}(R) = \{r \in R \mid r \text{ is nilpotent}\}.$$

(i) Prove that $\mathcal{N}(R)$ is an ideal of R. It is called the *nilradical* of R.

(ii) Prove that in $R/\mathcal{N}(R)$ the only nilpotent element is $0_{R/\mathcal{N}(R)}$.

3.19. For $n \in \mathbb{N}$, determine the number of nilpotent elements and idempotent elements in $\mathbb{Z}_n$. (These numbers depend on the prime factorization of n.) (Hint: Consider first the case where n is a prime power.)

3.20. Let R be a commutative ring, and let $f = \sum_{i=0}^{n} r_i X^i \in R[X]$ (with each $r_i \in R$).

(i) Prove that f is nilpotent iff each r_i is nilpotent.

(ii) Prove that f is a unit of $R[X]$ iff r_0 is a unit of R and $r_1, r_2, \ldots, r_n$ are each nilpotent. (Hint: Note that in any commutative ring R, if $u \in R^*$ and s is nilpotent then $u + s \in R^*$, as you can see from the geometric series expansion of $(1 + (u^{-1}s))^{-1}$.)

3.21. Consider the formal power series ring $R[[X]]$ for any ring R.

(i) Let $f = \sum_{i=0}^{\infty} a_i X^i \in R[[X]]$ with $a_0 = 1$. Prove that there is $g \in R[[X]]$ with $fg = 1$. (Hint: Write down a system of equations for the coefficients of g, and show that they can be solved recursively.) Of course, the analogous argument shows that there is $h \in R[[X]]$ with $hf = 1$. Then, $h = h1 = hfg = 1g = g$, so $f \in R[[X]]^*$.

(ii) Prove that
$$R[[X]]^* = \Big\{ \sum_{i=0}^{\infty} b_i X^i \mid b_0 \in R^* \Big\}.$$

Characteristic of a ring. Let R be a ring. Consider the map
$$\chi\colon \mathbb{Z} \to R \quad \text{given by} \quad \chi(j) = j1_R,$$
which is the unique ring homomorphism from $\mathbb{Z}$ to R. (It is unique because a homomorphism from $\mathbb{Z}$ is determined by the image of $1_{\mathbb{Z}}$, and a ring homomorphism is required to map $1_{\mathbb{Z}}$ to 1_R.) Its kernel is an ideal of $\mathbb{Z}$, hence a principal ideal by Example 3.1. The *characteristic* of R is
$$\mathit{char}(R) = \text{the unique integer } k \geq 0 \text{ such that } \mathit{ker}(\chi) = k\mathbb{Z}. \quad (3.25)$$
Note that for any integer multiple ℓ of $\mathit{char}(R)$ and any $r \in R$, we have $\ell r = (\ell 1_R)r = 0_R$. The *prime subring* of R is
$$P_R = \mathit{im}(\chi) = \{j1_R \mid j \in \mathbb{Z}\}. \quad (3.26)$$
This P_R is a subring of R, since it is the image of a ring homomorphism, and it lies in every other subring of R. Note that
$$\mathit{char}(R) = 0 \quad \text{iff} \quad P_R \cong \mathbb{Z}. \quad (3.27)$$
Suppose now that $\mathit{char}(R) \geq 1$. Then, as
$$P_R = \mathit{im}(\chi) \cong \mathbb{Z}/\mathit{ker}(\chi),$$
we have
$$\mathit{char}(R) = k > 0 \quad \text{iff} \quad P_R \cong \mathbb{Z}_k. \quad (3.28)$$

Note that for any subring T of R, we have $P_T = P_R$ since $1_T = 1_R$; hence, $char(T) = char(R)$.

3.22. Let R be any ring, and let $r, s \in R$ with $rs = sr$.

(i) Prove the binomial formula: For any $n \in \mathbb{N}$,

$$(r+s)^n = \sum_{i=0}^{n} \binom{n}{i} r^i s^{n-i}. \tag{3.29}$$

(ii) Suppose that $char(R) = p$, where p is a prime number. Prove that $(r+s)^p = r^p + s^p$ (recall equation (1.16)); hence, by induction,

$$(r+s)^{p^n} = r^{p^n} + s^{p^n}, \quad \text{for all } n \in \mathbb{N}. \tag{3.30}$$

3.23. *Construction of* $\mathbb{R}$ *from* $\mathbb{Q}$. There are a number of ways of building the real number system $\mathbb{R}$ starting from the rational numbers $\mathbb{Q}$, including using Dedekind cuts, or taking the completion of $\mathbb{Q}$ with respect to its Euclidean metric. This problem and the next two give a ring-theoretic approach to constructing $\mathbb{R}$ as the factor ring of the ring of Cauchy sequences of rational numbers modulo its ideal of null sequences.

(i) Consider the ring

$$T = \prod_{i=1}^{\infty} \mathbb{Q},$$

the direct product of countably infinitely many copies of $\mathbb{Q}$. An element $\alpha = (a_1, a_2, \ldots) \in T$ is said to be a *Cauchy sequence* if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that for all $i, j \in \mathbb{N}$, if $i > m$ and $j > m$ then $|a_i - a_j| < 1/n$. Let $\mathcal{C}$ be the set of all Cauchy sequences in T. Prove that $\mathcal{C}$ is a subring of T.

(ii) An element $\alpha = (a_1, a_2, \ldots) \in T$ is said to be a *null sequence* if for every $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that for $i \in \mathbb{N}$, if $i > m$ then $|a_i| < 1/n$ (i.e., the sequence converges to 0). Let $\mathcal{N}$ be the set of all null sequences in T. Prove that $\mathcal{N}$ is an ideal of $\mathcal{C}$.

(iii) Suppose $\alpha = (a_1, a_2, \ldots) \in \mathcal{C} \setminus \mathcal{N}$. Prove that there exist $k \in \mathbb{N}$ and $b \in \mathbb{Q}$ with $b > 0$ such that either (a) $a_i > b$ for every $i > k$; or (b) $a_i < -b$ for every $i > k$.

(iv) Let

$$\mathcal{R} = \mathcal{C}/\mathcal{N}.$$

Prove that $\mathcal{R}$ is a field. This is our candidate for $\mathbb{R}$, and the next two problems show that $\mathcal{R}$ has all the basic properties that characterize $\mathbb{R}$.

3.24. Let $\mathcal{R}$ be the field constructed in the previous problem. We now construct a total order on $\mathcal{R}$, so as to have inequalities and absolute values. This is done by first defining the positive elements. For a Cauchy sequence $\alpha = (a_0, a_1, \ldots) \in \mathcal{C}$, let $\overline{\alpha} = \alpha + \mathcal{N}$, which is the image of α in $\mathcal{R}$.

(i) Let

$$P = \left\{ \alpha = (a_0, a_1, \ldots) \in \mathcal{C} \;\middle|\; \begin{array}{l} \alpha \notin \mathcal{N}, \text{ and } a_i > 0 \\ \text{for all but finitely many } i \end{array} \right\}.$$

Prove that if $\alpha \in P$ and $\gamma \in \mathcal{N}$, then $\alpha + \gamma \in P$. Prove further that if $\alpha, \beta \in P$ then $\alpha + \beta \in P$ and $\alpha\beta \in P$.

(ii) Let $-P = \{-\alpha \mid \alpha \in P\}$. Prove that

$$\mathcal{C} = P \cup \mathcal{N} \cup -P,$$

a disjoint union.

(iii) Let $\mathcal{P} = \{\overline{\alpha} \mid \alpha \in P\} \subseteq \mathcal{R}$, which is the image of P in $\mathcal{R}$. Let $-\mathcal{P} = \{-\overline{\alpha} \mid \overline{\alpha} \in \mathcal{P}\}$. Prove that if $\overline{\alpha} \in \mathcal{P}$ and $\overline{\beta} \in \mathcal{P}$, then $\overline{\alpha} + \overline{\beta} \in \mathcal{P}$ and $\overline{\alpha}\overline{\beta} \in \mathcal{P}$. Prove further that

$$\mathcal{R} = \mathcal{P} \cup \{0_{\mathcal{R}}\} \cup -\mathcal{P},$$

a disjoint union.

(iv) Now define the relation $<$ on $\mathcal{R}$ by:

$$\overline{\alpha} < \overline{\beta} \quad \text{just when} \quad \overline{\beta} - \overline{\alpha} \in \mathcal{P}.$$

(Thus, $\mathcal{P} = \{\overline{\alpha} \in \mathcal{R} \mid 0 < \overline{\alpha}\}$.) Prove the "tricohtomy property" for $<$: for any $\overline{\alpha}, \overline{\beta} \in \mathcal{R}$, one and only one of the following holds: $\overline{\alpha} < \overline{\beta}$, or $\overline{\alpha} = \overline{\beta}$, or $\overline{\beta} < \overline{\alpha}$. Note further that if $\overline{\alpha} < \overline{\beta}$ and $\overline{\beta} < \overline{\gamma}$, then $\overline{\alpha} < \overline{\gamma}$. Additionally, if $\overline{\alpha} < \overline{\beta}$ and $\overline{\gamma} < \overline{\delta}$, then $\overline{\alpha} + \overline{\gamma} < \overline{\beta} + \overline{\delta}$; and, if $\overline{\alpha} < \overline{\beta}$ and $0 < \overline{\delta}$, then $\overline{\alpha}\overline{\delta} < \overline{\beta}\,\overline{\delta}$.

(v) Take any $\overline{\alpha}, \overline{\beta} \in \mathcal{R}$. We say that $\overline{\alpha} \le \overline{\beta}$ if $\overline{\alpha} < \overline{\beta}$ or $\overline{\alpha} = \overline{\beta}$. Define the absolute value of $\overline{\alpha}$ as usual:

$$|\overline{\alpha}| = \begin{cases} \overline{\alpha}, & \text{if } 0 \le \overline{\alpha}; \\ -\overline{\alpha}, & \text{if } \overline{\alpha} < 0. \end{cases}$$

Thus, we always have $0 \le |\overline{\alpha}|$; moreover $|\overline{\alpha}| = 0$ iff $\overline{\alpha} = 0$. Note also that $|\overline{\alpha}\,\overline{\beta}| = |\overline{\alpha}|\,|\overline{\beta}|$. Prove the "triangle inequality," that

$$|\overline{\alpha} + \overline{\beta}| \;\le\; |\overline{\alpha}| + |\overline{\beta}|.$$

(vi) Observe that there is a ring homomorphism $\iota\colon \mathbb{Q} \to \mathcal{R}$ given by $\iota(c) = \overline{(c, c, \ldots, c, \ldots)}$. Since $ker(\iota)$ is an ideal of the field $\mathbb{Q}$ and $1_{\mathbb{Q}} \notin ker(\iota)$ it follows that $ker(\iota) = \{0_{\mathbb{Q}}\}$, hence ι is injective. Prove that for $c, d \in \mathbb{Q}$,

$$c < d \text{ in } \mathbb{Q} \quad \text{iff} \quad \iota(c) < \iota(d) \text{ in } \mathcal{R}.$$

Thus, when we view $\mathbb{Q}$ as a subfield of $\mathcal{R}$ by identifying $\mathbb{Q}$ with its isomorphic copy $\iota(\mathbb{Q})$, the ordering defined on $\mathcal{R}$ restricts to the usual ordering on $\mathbb{Q}$.

3.25. Let $\mathcal{R}$ be the field of the previous two problems. This problem shows that $\mathcal{R}$ satisfies the fundamental properties of the real numbers, hence justifying defining $\mathbb{R} = \mathcal{R}$. Note that since we have a well-behaved absolute value on $\mathcal{R}$, we can define limits and Cauchy sequences of elements of $\mathcal{R}$: For an infinite sequence $\overline{\alpha_1}, \overline{\alpha_2}, \ldots$ of elements of $\mathcal{R}$ and any $\overline{\beta} \in \mathcal{R}$ we say that

$$\lim_{i\to\infty} \overline{\alpha_i} = \overline{\beta} \quad \text{if} \quad \begin{cases} \text{for every } n \in \mathbb{N}, \text{ there is } m \in \mathbb{N} \text{ such} \\ \text{that } \left|\overline{\beta} - \overline{\alpha_k}\right| < 1/n \text{ for every } k > m. \end{cases}$$

When this occurs, we say that the sequence of $\overline{\alpha_i}$ *converges* to $\overline{\beta}$. Cauchy sequences in $\mathcal{R}$ are defined as in problem 3.23(i), with elements of $\mathbb{Q}$ replaced by elements of $\mathcal{R}$.

(i) Prove the completeness property of $\mathcal{R}$: Every Cauchy sequence of elements of $\mathcal{R}$ converges to some element of $\mathcal{R}$.

(ii) Prove that $\mathbb{Q}$ is dense in $\mathcal{R}$, i.e., that for every $\overline{\alpha}$ in $\mathcal{R}$ there is a sequence of rational numbers that converges to $\overline{\alpha}$.

(iii) Let A and B be nonempty subsets of $\mathcal{R}$ such that $A \cap B = \varnothing$, $A \cup B = \mathcal{R}$, and $\overline{\alpha} < \overline{\beta}$ for every $\overline{\alpha} \in A$ and $\overline{\beta} \in B$. Prove

that there exists $\overline{\gamma} \in \mathcal{R}$ such that $\overline{\alpha} \leq \overline{\gamma} \leq \overline{\beta}$ for all $\overline{\alpha} \in A$ and $\overline{\beta} \in B$. Prove also that there is at most one such $\overline{\gamma}$.

(iv) Prove the *least upper bound property* for $\mathcal{R}$: Let A be a nonempty subset of $\mathcal{R}$. An *upper bound* of A is a $\overline{\beta} \in \mathcal{R}$ such that $\overline{\alpha} \leq \overline{\beta}$ for every $\overline{\alpha} \in A$. Prove that if A has an upper bound, then it has a *least upper bound*, i.e., an upper bound $\overline{\gamma}$ of A such that $\overline{\gamma} \leq \overline{\beta}$ for every upper bound $\overline{\beta}$ of A.

3.3. Polynomial rings and evaluation maps

3.26. Prove the general *Division Algorithm* for polynomials, which says: Let R be any ring. Take any nonzero $f \in R[X]$ whose leading coefficient is a unit of R, and any $g \in R[X]$. Then there exist unique $q, h \in R[X]$ such that

$$g = qf + h, \quad \text{with } deg(h) < deg(f) \text{ or } h = 0. \tag{3.31}$$

(Likewise, there exist unique $q', h' \in R[X]$ such that $g = fq' + h'$ with $deg(h') < deg(f)$ or $h' = 0$.)

Evaluation homomorphism. Let T be a commutative ring, and let R be a subring of T. Fix any $t \in T$. For any polynomial $f = \sum_{i=0}^{k} a_i X^i \in R[X]$, the *evaluation of f at t* is defined to be

$$f(t) = \sum_{i=0}^{k} a_i t^i.$$

If $f(t) = 0$, we say that t is a *root* of f. Allowing f to vary, we get the *evaluation at t function*

$$\varepsilon_{R,t} \colon R[X] \longrightarrow T \quad \text{given by} \quad f \mapsto f(t). \tag{3.32}$$

(Usually, the ring R in question is clear, and we write ε_t for $\varepsilon_{t,R}$.) Note that ε_t is a ring homomorphism. Let

$$R[t] = im(\varepsilon_t) = \{f(t) \mid f \in R[X]\}. \tag{3.33}$$

Then $R[t]$ is a subring of T, since it is the image of a ring homomorphism; $R[t]$ is called *the subring of T generated by t over R*, and it clearly lies in any subring of T containing R and t. Note that

$$ker(\varepsilon_t) = \{f \in R[X] \mid t \text{ is a root of } f\},$$

which is an ideal of $R[X]$. By the FHT, we have the extremely useful ring isomorphism

$$R[t] \cong R[X]/\ker(\varepsilon_t). \tag{3.34}$$

Example 3.27. Let R be a commutative ring, and let $r \in R$. Take any $g \in R[X]$. By the Division Algorithm, $g = q \cdot (X - r) + h$ for $q, h \in R[X]$, with $\deg h = 0$ or $h = 0$. By evaluating the equation for g at r (i.e., applying ε_r), we obtain that $g(r) = q(r) \cdot 0 + h(r) = h$. Thus,

$$g = q \cdot (X - r) + g(r).$$

It follows that r is a root of g iff g is a multiple of $(X - r)$ in $R[X]$. Thus, $\ker(\varepsilon_r) = (X - r)R[X]$, while $\mathrm{im}(\varepsilon_r) = R[r] = R$. Thus, (3.34) yields

$$R[X]/(X - r) \cong R. \tag{3.35}$$

3.28.

(i) Let R be a commutative ring, and let I be a nontrivial ideal of $R[X]$. Let k be the least degree of nonzero elements of I. Suppose that there is $g \in I \setminus \{0\}$ such that $\deg(g) = k$ and the leading coefficient of g is a unit of R. Prove that $I = (g)$. (Hint: Apply the Division Algorithm.)

(ii) If F is a field, prove that every ideal of $F[X]$ is principal.

(iii) Let R be a subring of a commutative ring T, and let $t \in T$. Suppose that the ideal $\ker(\varepsilon_t)$ of $R[X]$ is nontrivial and that it contains a nonzero element g of least degree such that g has leading coefficient in R^*. So, $\ker(\varepsilon_t) = (g)$ by part (i). Let $k = \deg(g)$. Prove that every element s of $R[t]$ is expressible uniquely as

$$s = r_0 + r_1 t + \ldots + r_i t^i + \ldots + r_{k-1} t^{k-1}$$

with $r_0, r_1, \ldots, r_{k-1} \in R$. (Hint: Use the Division Algorithm.)

3.29. Let $s \in \mathbb{Q}$ and suppose $s^n \in \mathbb{Z}$ for some $n \in \mathbb{N}$. Prove that $s \in \mathbb{Z}$. (Use the prime factorization in $\mathbb{Z}$.)

3.30. Take any $d \in \mathbb{Z}$ such that $d \neq j^2$ for any $j \in \mathbb{Z}$. Let $\sqrt{d}$ denote either square root of d in $\mathbb{C}$. Note that $\sqrt{d} \notin \mathbb{Q}$ by the preceding problem. Consider the subrings $\mathbb{Q}[\sqrt{d}\,]$ and $\mathbb{Z}[\sqrt{d}\,]$ of $\mathbb{C}$.

(i) For the evaluation homomorphism $\varepsilon_{\mathbb{Q},\sqrt{d}}\colon \mathbb{Q}[X] \to \mathbb{C}$, prove that $ker(\varepsilon_{\mathbb{Q},\sqrt{d}}) = (X^2 - d)\mathbb{Q}[X]$. Deduce that

$$\mathbb{Q}[\sqrt{d}\,] \cong \mathbb{Q}[X]/(X^2 - d)\mathbb{Q}[X],$$

and that every element of $\mathbb{Q}[\sqrt{d}\,]$ is uniquely expressible as $r + s\sqrt{d}$ with $r, s \in \mathbb{Q}$. (See problem 3.28.) Prove also that $Q[\sqrt{d}\,]$ is a field. (Recall "rationalizing the denominator.")

(ii) For the evaluation homomorphism $\varepsilon_{\mathbb{Z},\sqrt{d}}\colon \mathbb{Z}[X] \to \mathbb{C}$ prove that $ker(\varepsilon_{\mathbb{Z},\sqrt{d}}) = (X^2 - d)\mathbb{Z}[X]$. Deduce that

$$\mathbb{Z}[\sqrt{d}\,] \cong \mathbb{Z}[X]/(X^2 - d)\mathbb{Z}[X]$$

and that every element of $\mathbb{Z}[\sqrt{d}\,]$ is uniquely expressible as $k + \ell\sqrt{d}$ with $k, \ell \in \mathbb{Z}$. Prove also that $\mathbb{Z}[\sqrt{d}\,]$ is not a field.

3.31. Let F be a field, let $f \in F[X]$ be nonzero, and let $k = deg(f)$. Prove that f has at most k roots in F.

3.32. Let F be a field, and let U be a finite subgroup of its multiplicative group F^*. Prove that U is a cyclic group. (Hint: Apply the preceding problem and problem 2.20.)

3.33. Let f be nonconstant in $\mathbb{Z}[X]$. Prove that there are infinitely many prime numbers p such that the image of f in $\mathbb{Z}_p[X]$ (obtained by reducing the coefficients of f modulo p as in Example 3.15(ii)) has a root in $\mathbb{Z}_p$.

Zero divisors. Let R be a commutative ring. A nonzero $r \in R$ is called a *zero divisor* if there is a nonzero $a \in R$ with $ra = 0$. Note that units of R are never zero divisors. The non-zero-divisors are the elements for which multiplicative cancellation is always possible: If $s \in R$ with $s \neq 0$, then s is not a zero divisor iff whenever $sa = sb$, we have $a = b$, for any $a, b \in R$.

3.34. Let R be a commutative ring with $|R| < \infty$. Prove that every nonzero element of R is either a unit or a zero divisor.

3.35. Let R be a commutative ring, and let $f \in R[X]$. Suppose there is $g \in R[X]$ with $g \neq 0$ but $gf = 0$. Prove that there is $r \in R$ with $r \neq 0$ and $rf = 0$. It follows that if f is a zero divisor in $R[X]$, then each nonzero coefficient of f is a zero divisor in R.

3.4. Integral domains, quotient fields

Integral domains. An *integral domain* is a commutative ring R such that $1_R \neq 0_R$ and for any $a, b \in R$ if $ab = 0$, then $a = 0$ or $b = 0$. Restated, an integral domain is a nontrivial commutative ring with no zero divisors. Thus, in an integral domain R we have the multiplicative cancellation property that is familiar for $\mathbb{Z}$: If $a, b, c \in R$ and $c \neq 0$, then

$$\text{if } \ ac = bc, \quad \text{then } \ a = b.$$

For example, any field is an integral domain, as is $\mathbb{Z}$. Also, $\mathbb{Z}_n$ is an integral domain iff n is a prime number. Note that any subring of an integral domain is also an integral domain. Also, if R is an integral domain, then for nonzero $f, g \in R[X]$ we have $fg \neq 0$ and

$$deg(fg) \ = \ deg(f) + deg(g).$$

Hence, $R[X]$ is also an integral domain. The degree formula shows that units of $R[X]$ must then have degree 0. Thus, $R[X]^* = R^*$. Note also that since R is an integral domain, its prime subring P_R is an integral domain as well; so $P_R \cong \mathbb{Z}$ or $P_R \cong \mathbb{Z}_p$ for some prime number p. (See (3.27) and (3.28).) Hence, $char(R) = 0$ or $char(R) = p$ for p prime.

3.36. Let F be a field, and let $\theta\colon F[X] \to F[X]$ be a (ring) automorphism. Prove that the restriction of θ to F is an automorphism of F, and that $\theta(X) = aX + b$, for some $a \in F^*$ and $b \in F$. (Note that the converse of this is clearly true: If σ is any automorphism of F and $a \in F^*$ and $b \in F$ then $\theta\colon F[X] \to F[X]$ given by

$$\theta\Big(\sum_{i=0}^{n} c_i X^i\Big) \ = \ \sum_{i=0}^{n} \sigma(c_i)(aX + b)^i$$

is an automorphism of $F[X]$.)

3.37. Give an example of an integral domain R and an automorphism $\theta\colon R[X] \to R[X]$, such that $deg(\theta(X)) \neq 1$.

Quotient fields. Let R be an integral domain. We recall the construction of the quotient field of R: Define a relation $\approx$ on the Cartesian product $R \times (R \setminus \{0\})$ by

$$(r, s) \approx (r', s') \ \text{ iff } \ rs' \ = \ r's.$$

It is easy to check that $\approx$ is an equivalence relation. For any (r, s) in $R \times (R \setminus \{0\})$, let r/s denote the equivalence class of (r, s),

$$r/s = \{(r', s') \in R \times (R \setminus \{0\}) \mid rs' = r's\}. \tag{3.36}$$

Thus,

$$r/s = r'/s' \quad \text{iff} \quad (r, s) \approx (r', s') \quad \text{iff} \quad rs' = r's.$$

Let $q(R)$ denote the set of equivalence classes:

$$q(R) = \{r/s \mid r \in R,\ s \in R \setminus \{0\}\}. \tag{3.37}$$

Define operations $+$ and $\cdot$ on $q(R)$ by

$$r/s + r'/s' = (rs' + r's)/(ss') \quad \text{and} \quad r/s \cdot r'/s' = (rr')/(ss').$$

Straightforward calculations show that these operations are well-defined (independent of the choice of representative (r, s) for r/s and (r', s') for r'/s'), and that with these operations $q(R)$ is a commutative ring. Note that

$$r/s = 0_{q(R)} \quad \text{iff} \quad r = 0,$$

and that

$$1_{q(R)} = s/s \quad \text{for any } s \in R \setminus \{0\}.$$

If $r/s \neq 0$, then $r \in R \setminus \{0\}$, and $r/s \cdot s/r = 1_{q(R)}$, so $r/s \in q(R)^*$, with $(r/s)^{-1} = s/r$. Thus, $q(R)$ is a field, called the *quotient field of* R. There is an injective ring homomorphism $\iota\colon R \to q(R)$ given by $\iota(r) = r/1$. Note that for any $r/s \in q(R)$, we have

$$r/s = \iota(r)\iota(s)^{-1},$$

justifying the fraction notation for elements of $q(R)$. We often view R as a subring of $q(R)$, by identifying R with its isomorphic copy $\iota(R)$.

Note that the quotient field construction generalizes the construction of $\mathbb{Q}$ from $\mathbb{Z}$. Thus, $\mathbb{Q} = q(\mathbb{Z})$.

3.38. Let K be a field, and let R be a subring of K; so, R is an integral domain. Let $q(R)$ and $\iota\colon R \to q(R)$ be as described above. Prove that there is a well-defined injective ring homomorphism $\beta\colon q(R) \to K$ given by

$$\beta(r/s) = rs^{-1} \quad \text{for all } (r, s) \in R \times (R \setminus \{0\}).$$

Prove also that β is the unique ring homomorphism from $q(R)$ to K such that $\beta(\iota(r)) = r$ for all $r \in R$. Note that since β is injective,

$$q(R) \cong \mathit{im}(\beta) = \{rs^{-1} \mid r \in R,\ s \in R \setminus \{0\}\} \subseteq K.$$

The field $im(\beta)$ is the subfield of K generated by R since it clearly contains R and lies in every subfield of K containing R. Because of the isomorphism $q(R) \cong im(\beta)$, $im(\beta)$ is informally called "the quotient field of R in K."

3.39. Let $R \subseteq T$ be integral domains. View T as a subring of $q(T)$. Suppose that for every $t \in T$ there is $r \in R$ with $r \neq 0$ such that $rt \in R$. Prove that the quotient field of R in $q(T)$ is $q(T)$. Thus, $q(R) \cong q(T)$, and informally we say that "R and T have the same quotient field."

3.40. Let $R \subseteq T$ be integral domains. View T and $q(R)$ as subrings of $q(T)$. Take any $t \in T$. Prove that $R[t]$ and $q(R)[t]$ have the same quotient field in $q(T)$.

Example 3.41. Here are examples where the two preceding problems apply:

(i) Let R be an integral domain. Then $R[X]$ and $q(R)[X]$ have the same quotient field.

(ii) Let R be an integral domain, and let T be a subring of $q(R)$ with $R \subseteq T$. Then, R and T have the same quotient field.

(iii) Let $d \in \mathbb{Z}$ with $d \neq j^2$ for any $j \in \mathbb{Z}$, and let $\sqrt{d}$ be either square root of d in $\mathbb{C}$. Since $\mathbb{Q} = q(\mathbb{Z})$, the integral domains $\mathbb{Z}[\sqrt{d}\,]$ and $\mathbb{Q}[\sqrt{d}\,]$ have the same quotient field. Since $\mathbb{Q}[\sqrt{d}\,]$ is a field (recall problem 3.30), it is its own quotient field. Thus, $\mathbb{Q}[\sqrt{d}\,]$ is the quotient field of $\mathbb{Z}[\sqrt{d}\,]$.

3.42. Consider the formal power series ring $F[[X]]$ over a field F. Recall the description of $F[[X]]^*$ given in problem 3.21.

(i) Take any nonzero $f = \sum_{i=0}^{\infty} a_i X^i \in F[[X]]$, and let j be minimal such that $a_j \neq 0$. Prove the equality of principal ideals $(f) = (X^j)$.

(ii) Deduce that the only ideals of $F[[X]]$ are $\{0\}$ and (X^i) for $i = 0, 1, 2, \ldots$.

(iii) View $F[[X]]$ as a subring of its quotient field $q(F[[X]])$. Let $F[[X]][X^{-1}]$ be the subring of $q(F[[X]])$ generated by $F[[X]]$ and X^{-1}. Prove that $F[[X]][X^{-1}]$ is a field; thus,

$$q(F[[X]]) \;=\; F[[X]][X^{-1}].$$

Note that the elements of $F[[X]][X^{-1}]$ are conveniently expressible as formal *Laurent series*, i.e., in the form $\sum_{i=k}^{\infty} a_i X^i$ for some k in $\mathbb{Z}$, with all $a_i \in F$.

3.5. Maximal ideals and prime ideals

Maximal ideals. An ideal M of a ring R is a *maximal ideal* if it is maximal among proper ideals of R, i.e., $M \neq R$ and there is no ideal J of R with $M \subsetneqq J \subsetneqq R$. Note that if R is commutative, then an ideal M is a maximal ideal of R iff R/M is a field. This follows from the Second Isomorphism Theorem and the characterization of fields in terms of ideals. For example, in $\mathbb{Z}$, the maximal ideals are the ideals $p\mathbb{Z}$ for p a prime number.

3.43. Let I be a proper ideal of a ring R. Use Zorn's Lemma to prove that there is a maximal ideal M of R with $I \subseteq M$. (Hint: Consider the set of ideals J of R with $I \subseteq J$ and $1 \notin J$.)

3.44. Let F be a field, and let

$$T = \left\{ \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right) \mid a, b, c \in F \right\},$$

the subring of upper triangular matrices in $M_2(F)$, and let

$$J = \left\{ \left(\begin{smallmatrix} 0 & b \\ 0 & 0 \end{smallmatrix}\right) \mid b \in F \right\}.$$

(i) Prove that J is an ideal of T with $J^2 = \{0\}$, and that

$$T/J \cong F \times F.$$

(ii) Prove that J lies in every maximal ideal of T.

(iii) Determine the maximal ideals of T/J, and use this to determine the maximal ideals of T.

3.45. Let $[0,1] = \{r \in \mathbb{R} \mid 0 \leq r \leq 1\}$, the closed unit interval in $\mathbb{R}$, and let $C[0,1]$ be the set of continuous functions $f\colon [0,1] \to \mathbb{R}$. Note that $C[0,1]$ is a commutative ring with pointwise operations as follows: for $f, g \in C[0,1]$ define $f+g$ and $f \cdot g$ by

$$(f+g)(r) = f(r) + g(r) \quad \text{and} \quad (f \cdot g)(r) = f(r)g(r),$$

for all $r \in [0,1]$. So, $0_{C[0,1]}$ is the constant function $r \mapsto 0$ for all $r \in [0,1]$, and $1_{C[0,1]}$ is the constant function $r \mapsto 1$.

(i) Prove that

$$C[0,1]^* = \{f \in C[0,1] \mid f(r) \neq 0 \text{ for every } r \in [0,1]\}.$$

(ii) Fix $c \in [0,1]$, and let

$$M_c = \{f \in C[0,1] \mid f(c) = 0\}.$$

Prove that M_c is a maximal ideal of $C[0,1]$ with

$$C[0,1]/M_c \cong \mathbb{R}.$$

(iii) Prove that every proper ideal of $C[0,1]$ lies in M_c for some $c \in [0,1]$. (Use the compactness of $[0,1]$.) Hence, every maximal ideal of $C[0,1]$ is one of the M_c.

(iv) Let $(0,1] = \{r \in \mathbb{R} \mid 0 < r \leq 1\}$ and, as above, let $C(0,1]$ be the ring of continuous functions $(0,1] \to \mathbb{R}$. For any $c \in (0,1]$, the ideal $M'_c = \{f \in C(0,1] \mid f(c) = 0\}$ is a maximal ideal of $C(0,1]$. But, find a proper ideal of $C(0,1]$ that does not lie in any M'_c. It follows that there are maximal ideals of $C(0,1]$ other than the M'_c.

Prime ideals. Let R be a commutative ring. An ideal P of R is called a *prime ideal* if $P \neq R$ and for all $a, b \in R$, if $ab \in P$, then $a \in P$ or $b \in P$. Note that P is prime ideal of R iff R/P is an integral domain. Thus, the FHT shows that if $f\colon R \to T$ is any ring homomorphism, then $im(f)$ is an integral domain iff $ker(f)$ is a prime ideal of R. Note also that every maximal ideal M of R is a prime ideal, since R/M is a field and hence an integral domain. For example, the prime ideals of $\mathbb{Z}$ are $\{0\}$, which is not maximal, and the maximal ideals (p) where p is a prime number.

Note that if I and P are ideals of a commutative ring R with $P \supseteq I$, then P is a prime ideal of R iff P/I is a prime ideal of R/I. This follows immediately from the isomorphism $R/I \,/\, P/I \cong R/P$ given by the Second Isomorphism Theorem. Note also that if $R \subseteq T$ are commutative rings and Q is a prime ideal of T, then $Q \cap R$ is a prime ideal of R.

3.46. Let R_1 and R_2 be commutative rings. Recall (see problem 3.7(i)) that every ideal of $R_1 \times R_2$ has the form $I_1 \times I_2$, where each I_i is an ideal of R_i. Prove that every prime ideal of $R_1 \times R_2$ has the form

$P_1 \times R_2$ or $R_1 \times P_2$, where P_i is a prime ideal of R_i. Likewise, prove that every maximal ideal of $R_1 \times R_2$ has the form $M_1 \times R_2$ or $R_1 \times M_2$, where M_i is a maximal ideal of R_i.

3.47. Let R be a finite commutative ring.

(i) Prove that if R is an integral domain, then R is a field.

(ii) Prove that every prime ideal of R is a maximal ideal.

3.48. Let R be a commutative ring, and fix $s \in R$. Let

$$T = R[X]/(1 - sX).$$

Let $\psi\colon R \to T$ be the composition of the standard homomorphisms

$$R \longrightarrow R[X] \longrightarrow R[X]/(1 - sX) = T,$$

i.e., $\psi(r) = r + (1 - sX)$. Let $R' = im(\psi)$, the image of R in T. We can think of T as obtained by "enlarging" R by adjoining a multiplicative inverse for s. (Clearly, $T = R'[x]$, where $x = X + (1 - sX)$ is the inverse of $s' = \psi(s)$ in T.) However, there is an obstruction to obtaining a ring containing R and an inverse for s: If s is a zero divisor in R or if $s = 0$ and R is nontrivial, then s cannot be a unit in R, nor in any ring containing R. This obstacle is dealt with by passing from R to R', in which the image of s is not a zero divisor, then enlarging R' by adjoining an inverse of s':

(i) Let

$$J = \{r \in R \mid s^n r = 0 \text{ for some } n \in \mathbb{N}\}.$$

Prove that

$$J = (1 - sX) \cap R = ker(\psi).$$

Thus, $R' \cong R/J$.

(ii) Prove (without using T) that $s + J$ is not a zero divisor in R/J.

(iii) Let I be an ideal of R such that $s + I$ is not a zero divisor in R/I. Prove that $I \supseteq J$.

(iv) Prove that T is a trivial ring iff s is nilpotent in R.

(v) Prove that if R is an integral domain and $s \neq 0$, then $T \cong R[1/s]$, the subring of $q(R)$ generated by R and $1/s$.

(For more about T when R is a UFD, see problem 3.75 below.)

(vi) Prove that there is a one-to-one correspondence between the prime ideals Q of T and those prime ideals P of R with $s \notin P$. The correspondence is given by $Q \mapsto \psi^{-1}(Q)$ and $P \mapsto \psi(P)T$, the ideal of T generated by $\psi(P)$.

3.49. Let R be a commutative ring. Recall from problem 3.18 that the nilradical of R is

$$\mathcal{N}(R) = \{r \in R \mid r \text{ is nilpotent}\}.$$

Prove that

$$\mathcal{N}(R) = \bigcap_{P \text{ prime}} P, \tag{3.38}$$

the intersection of all the prime ideals of R. (Hint: Use the preceding problem and the fact that for $s \in R$, if $R[X]/(1 - sX)$ is nontrivial, then by problem 3.43 it has a maximal ideal.)

Polynomials in more than one variable. Let R be any ring. We have previously considered the polynomial ring $R[X]$. Let Y be a new indeterminate different from X. Then, the *polynomial ring over R in X and Y*, denoted $R[X, Y]$, is the iterated polynomial ring

$$R[X, Y] = \big(R[X]\big)[Y]. \tag{3.39}$$

Thus,

$$R[X, Y] = \Big\{ \sum_{i=0}^{m} \sum_{j=0}^{n} r_{ij} X^i Y^j \mid m, n \in \mathbb{N}, \text{ and all } r_{ij} \in R \Big\},$$

with

$$\sum_{i=0}^{m} \sum_{j=0}^{n} r_{ij} X^i Y^j = \sum_{i=0}^{m} \sum_{j=0}^{n} s_{ij} X^i Y^j \quad \text{iff} \quad r_{ij} = s_{ij} \text{ for all } i, j,$$

and with the usual rules for adding and multiplying polynomials. Note that $YX = XY$ in $R[X, Y]$. Also, there is a canonical isomorphism $R[X, Y] \to R[Y, X]$ given by $\sum_i \sum_j r_{ij} X^i Y^j \mapsto \sum_i \sum_j r_{ij} Y^j X^i$; hence, we will identify $R[Y, X]$ with $R[X, Y]$. Note further that if $R \subseteq T$ are commutative rings and $s, t \in T$, then there is a well-defined evaluation ring homomorphism $\varepsilon_{R,s,t} \colon R[X, Y] \to T$ given by: if $f = \sum_{i=0}^{m} \sum_{j=0}^{n} r_{ij} X^i Y^j \in R[X, Y]$, then

$$\varepsilon_{R,s,t}(f) = f(s, t) = \sum_{i=0}^{m} \sum_{j=0}^{n} r_{ij} s^i t^j \tag{3.40}$$

Thus, $im(\varepsilon_{R,s,t}) = R[s,t]$, the subring of T generated by R, s, and t.

In a similar manner, we construct polynomials in more than two variables: Let $X_1, \ldots, X_n$ be n distinct indeterminates for $n \geq 2$, and define recursively

$$R[X_1, \ldots, X_n] = \big(R[X_1, \ldots, X_{n-1}]\big)[X_n]. \tag{3.41}$$

All the remarks made above about polynomials in two variables have obvious analogues for polynomials in more than two variables.

3.50. Let R be a commutative ring, and let $r_1, \ldots, r_n \in R$. Prove that the evaluation homomorphism $\varepsilon_{R,r_1,\ldots,r_n} \colon R[X_1, \ldots, X_n] \to R$ has kernel $(X_1 - r_1, \ldots, X_n - r_n)$, the ideal of $R[X_1, \ldots, X_n]$ generated by the $X_i - r_i$.

3.51. Let R be in integral domain with $|R| = \infty$, and take any $f_1, \ldots, f_k$ in $R[X_1, \ldots, X_n]$. Suppose that for every $r_1, \ldots, r_n \in R$ there is an i with $f_i(r_1, \ldots, r_n) = 0$. Prove that $f_j = 0$ for some j.

Note 3.52. The $k = 2$ case of the preceding problem can be restated as follows: Let R be an infinite integral domain, and take any $f, g \in R[X_1, \ldots, X_n]$ with $g \neq 0$. For $r_1, \ldots, r_n \in R$,

$$\text{if } f(r_1, \ldots, r_n) = 0 \text{ whenever } g(r_1, \ldots, r_n) \neq 0, \text{ then } f = 0.$$

3.6. Divisibility and principal ideal domains

Irreducible elements. Let R be an integral domain. Recall that for $a, b \in R$ we say that a *divides* b (written $a|b$) if there is $c \in R$ with $b = ac$. So, $a|b$ iff $(b) \subseteq (a)$. We write

$$a \sim b \quad \text{if } a|b \text{ and } b|a; \tag{3.42}$$

a and b are then said to be *associates*. This occurs iff $(a) = (b)$. Since R is an integral domain, $a \sim b$ iff $b = ua$ (so $a = u^{-1}b$) for some $u \in R^*$. Every $b \in R$ has a *trivial factorization* $b = u^{-1}(ub)$ for any $u \in R^*$. A nonzero nonunit b of R is said to be *irreducible* if whenever $b = ac$ with $a, c \in R$ either a or c lies in R^*. Thus, b is irreducible iff b has no nontrivial factorization in R. Clearly, associates of irreducible elements of R are irreducible. The irreducibles are the building blocks in any analysis of factorization of elements of R.

Prime elements. A nonzero nonunit q of an integral domain R is said to be a *prime element* of R if

$$q|ab \quad \text{implies} \quad q|a \ \text{ or } \ q|b, \quad \text{for any } a, b \in R.$$

Thus, q is a prime element of R iff (q) is a nonzero prime ideal of R. Associates of prime elements of R are again prime elements. It is easy to check that if q is a prime element of R, then q is irreducible in R. However, irreducible elements need not be prime elements—see examples in problems 3.60 and 3.68 below.

3.53. Divisibility is defined for elements in any commutative ring R, but there are significant complications when R has zero divisors. Here is an example. Let $R = \mathbb{Z}[X]/(5X)$, which is not an integral domain. For $f \in \mathbb{Z}[X]$, let $\overline{f}$ denote the image $f + (5X)$ of f in R.

(i) Determine R^*.

(ii) Prove that $\overline{X} \,\big|\, \overline{2X}$ and $\overline{2X} \,\big|\, \overline{X}$ in R, but there is no $u \in R^*$ with $\overline{2X} = u\overline{X}$.

Greatest common divisors and least common multiples. Let R be an integral domain, and take any $a, b \in R$. An element d of R is called a *greatest common divisor* of a and b if both

(i) $d|a$ and $d|b$; and

(ii) for any $e \in R$, if $e|a$ and $e|b$, then $e|d$.

If d is a greatest common divisor of a and b, then clearly so is ud for any $u \in R^*$. Moreover, if d' is another greatest common divisor of a and b, then $d|d'$ and $d'|d$. Thus, greatest common divisors (when they exist) are uniquely determined up to associates in R. We write

$$d \sim gcd(a, b)$$

to indicate that d is a greatest common divisor of a and b. Note that for any $c \in R \setminus \{0\}$, we have

$$d \sim gcd(a, b) \ \text{ iff } \ cd \sim gcd(ca, cb). \tag{3.43}$$

Greatest common divisors of more than two elements are defined analogously.

Least common multiples are defined similarly, with the divisibility relations reversed: For elements a and b of integral domain R, an element ℓ of R is a *least common multiple* of a and b if both

(i) $a|\ell$ and $b|\ell$; and

(ii) for any $e \in R$, if $a|e$ and $b|e$, then $\ell|e$.

Note that least common multiples (when they exist) are unique up to associates in R. We write

$$\ell \sim lcm(a, b)$$

when ℓ is a least common multiple of a and b.

An integral domain R is called a *principal ideal domain* (abbreviated PID) if every ideal of R is a principal ideal. For example, $\mathbb{Z}$ is a PID, as is F, $F[X]$, and $F[[X]]$ for any field F (see Example 3.1 and problems 3.28(ii) and 3.42(ii)). See below just before problem 3.61 and problem 3.73 for further examples of PID's. The next problem shows that $\mathbb{Z}[X]$ and $F[X, Y]$ are not PID's.

3.54. Let R be a commutative ring, and let $r \in R$ with $r \notin R^*$ and r not a zero divisor. Prove that the ideal (r, X) of the polynomial ring $R[X]$ is not a principal ideal. (It follows that if $R[X]$ is a PID, then R must be a field.)

3.55. Let R be a commutative ring.

(i) Take $r \in R$ with $r \notin R^*$ and $r \neq 0$ such that the ideal (r, X) of $R[X]$ is principal. Prove that there is an idempotent element e of R with $e \neq 0$, $e \neq 1$ such that $r = re$ and r is a unit of the ring Re.

(ii) Now prove the converse: If $R = R_1 \times R_2$, a direct product of commutative rings, and $u \in R_1^*$, prove that in $R[X]$,

$$\big((u, 0), X\big) = \big((0, 1)X + (1, 0)\big).$$

3.56. Let R be an integral domain, and take any $a, b \in R$.

(i) Prove that if the ideal (a, b) of R generated by a and b is a principal ideal, say $(a, b) = (c)$, then c is a *gcd* of a and b such that $c = ra + sb$ for some $r, s \in R$.

(ii) Prove that if d is a *gcd* of a and b and $d = ta + ub$ for some $t, u \in R$, then $(a, b) = (d)$.

Thus, if R is a PID, any two elements have a greatest common divisor. For $R = \mathbb{Z}$, or $R = F[X]$ for F a field, or $R = \mathbb{Z}[\sqrt{-1}\,]$ (see (3.47) below) there is a division process that allows one to compute gcd's of elements of R by repeated long divisions, as in the Euclidean Algorithm.)

3.57. Let R be an integral domain, and let $r \in R$ with $r \notin R^*$ and $r \neq 0$.

(i) Prove that r is irreducible in R iff the principal ideal (r) is maximal among proper principal ideals of R, i.e., there is no principal ideal (s) of R with $(r) \subsetneqq (s) \subsetneqq R$.

(ii) Suppose R is a PID. Prove that the following conditions are equivalent for r as above:
   (a) r is irreducible in R.
   (b) (r) is a maximal ideal of R.
   (c) r is a prime element of R.

3.58. This problem is a straighforward application of the Second Isomorphism Theorem. The resulting isomorphism (3.44) is an extremely useful tool in analyzing factor rings, as illustrated in subsequent problems. Let R be a commutative ring, and let $a, b \in R$. Let $\overline{a} = a + (b)$ be the image of a in $R/(b)$.

(i) Prove that $\overline{a}\,R/(b) = (a, b)/(b)$ as ideals of $R/(b)$. ($\overline{a}\,R/(b)$ means $\overline{a}(R/(b))$, the principal ideal of $R/(b)$ generated by $\overline{a}$.)

(ii) Prove that $R/(b)\big/\big(\overline{a}\,R/(b)\big) \cong R/(a, b)$.

(iii) Let $\overline{a} = a + (b)$ as above, and let $\widetilde{b} = b + (a)$, which is the image of b in $R/(a)$. Prove that

$$R/(b)\big/\big(\overline{a}\,R/(b)\big) \;\cong\; R/(a)\big/\big(\widetilde{b}\,R/(a)\big). \tag{3.44}$$

3.59. Take $d \in \mathbb{Z}$ with $d \neq a^2$ for any $a \in \mathbb{Z}$, and let $\sqrt{d}$ be either square root of d in $\mathbb{C}$. Let $R = \mathbb{Z}[\sqrt{d}\,]$, a subring of $\mathbb{C}$. Recall from problem 3.30 that $R \cong \mathbb{Z}[X]/(X^2 - d)$. For any $n \in \mathbb{N}$, prove that

$$R/nR \;\cong\; \mathbb{Z}[X]/(n, X^2 - d) \;\cong\; \mathbb{Z}_n[X]/(X^2 - [d]_n) \tag{3.45}$$

(Hint: Apply the preceding problem.)

3.60. Let $d \in \mathbb{Z}$ with $d < 0$, let $\sqrt{d}$ be either square root of d in $\mathbb{C}$, and let $R = \mathbb{Z}[\sqrt{d}\,]$ a subring of $\mathbb{C}$. We know from problem 3.30(ii) that every element of R is uniquely expressible as $a + b\sqrt{d}$ with $a, b \in \mathbb{Z}$. The *norm map*, $N\colon \mathbb{Z}[\sqrt{d}\,] \to \mathbb{Z}$ is defined by

$$N(a + b\sqrt{d}\,) \;=\; (a + b\sqrt{d}\,)(a - b\sqrt{d}\,) \;=\; a^2 - b^2 d, \tag{3.46}$$

for all $a, b \in \mathbb{Z}$. Note that the norm map is multiplicative, i.e.,

$$N(rs) \;=\; N(r)N(s)$$

for all $r, s \in R$. Indeed, since $d < 0$, $N(r) = |r|^2$, where $|r|$ is the absolute value r as a complex number.

(i) For $r \in R$ prove that

$$r \in R^* \quad \text{iff} \quad N(r) \in \mathbb{Z}^* \;=\; \{\pm 1\}.$$

Deduce that

$$Z[\sqrt{-1}\,]^* \;=\; \{\pm 1, \pm\sqrt{-1}\,\},$$

while

$$Z[\sqrt{d}\,]^* \;=\; \{\pm 1\} \quad \text{if} \quad d \le -2.$$

(ii) Use the norm map to prove that every nonzero nonunit of R is a product of (one or more) irreducible elements.

(iii) If $d \le -2$, prove that 2 is irreducible in R.

(iv) Suppose that $d \le -3$. Prove that if d is even, then $2 \,\big|\, (\sqrt{d}\,)^2$ in R, but $2 \nmid \sqrt{d}$. Similarly, if d is odd, prove that $2 \big| (1+\sqrt{d}\,)^2$ but $2 \nmid (1 + \sqrt{d}\,)$. In either case, it follows that 2 is not a prime element of R.

(v) Suppose that $d \le -3$. Since R contains the irreducible element 2 that is not a prime element, R cannot be a PID. (See problem 3.57(ii).) To see this more specifically, show that if d is even, then the ideal $(2, \sqrt{d}\,)$ is not a principal ideal of R. (Hint: Since 2 is irreducible, if $(2, \sqrt{d}\,)$ is a principal ideal, then it equals R by problem 3.57(i).) Similarly, if d is odd, prove that $(2, 1 + \sqrt{d}\,)$ is not a principal ideal.

(vi) We now describe the prime ideals of $R = \mathbb{Z}[\sqrt{d}\,]$. This applies for any integer $d < 0$. Since R is an integral domain, $\{0\}$ is a prime ideal of R. Now, take any prime ideal P of R with $P \neq \{0\}$. Then, $P \cap \mathbb{Z}$ is a prime ideal of $\mathbb{Z}$. Moreover,

$P \cap \mathbb{Z} \neq \{0\}$, since if $r \in P \setminus \{0\}$ then $N(r) \in (P \cap \mathbb{Z}) \setminus \{0\}$. Hence, $P \cap \mathbb{Z} = p\mathbb{Z}$ for some prime number p. So, $P \supseteq pR$. Since an ideal I of R containing p is a prime ideal of R iff I/pR is a prime ideal of R/pR (as $R/I \cong R/pR/I/pR$), we need to understand the prime ideals of R/pR. For this we can use the isomorphism of (3.45) There are four cases:

(a) If $p|d$, then

$$\mathbb{Z}_p[X]/(X^2 - [d]_p) = \mathbb{Z}_p[X]/(X^2).$$

Prove that $P = (p, \sqrt{d}\,)$ is a prime ideal of R, and is the unique prime ideal of R with $P \cap R = p\mathbb{Z}$.

(b) Similarly, if $p = 2$ and d is odd, we have

$$\mathbb{Z}_p[X]/(X^2 - [d]_p) = \mathbb{Z}_2[X]/((X - [1]_2)^2).$$

Prove that $P = (2, 1 + \sqrt{d}\,)$ is the unique prime ideal of R with $P \cap \mathbb{Z} = 2\mathbb{Z}$.

(c) Let p be an odd prime with $p \nmid d$, and suppose there is $a \in \mathbb{Z}$ with $a^2 \equiv d \ (mod\ p)$. Then, in $\mathbb{Z}_p[X]$

$$X^2 - [d]_p = X^2 - [a]_p^2 = (X - [a]_p)(X + [a]_p),$$

and $[a]_p \neq -[a]_p$ as p is odd. Hence,

$$(X - [a]_p) + (X + [a]_p) = \mathbb{Z}_p[X].$$

The Chinese Remainder Theorem ((3.24) above) then shows that

$$\begin{aligned}\mathbb{Z}_p[X]/(X^2 - [d]_p) &= \mathbb{Z}_p[X] \,/\, (X - [a]_p)(X + [a]_p) \\ &\cong \mathbb{Z}_p[X]/(X - [a]_p) \times \mathbb{Z}_p[X]/(X + [a]_p) \\ &\cong \mathbb{Z}_p \times \mathbb{Z}_p.\end{aligned}$$

(Recall Example 3.27.) Prove that there are just two prime ideals, P of R with $P \cap \mathbb{Z} = p\mathbb{Z}$, and that they are $P_1 = (a - \sqrt{d}\,, p)$ and $P_2 = (-a - \sqrt{d}\,, p)$.

(d) Let p be an odd prime such that there is no $a \in \mathbb{Z}$ with $a^2 \equiv d \ (mod\ p)$. Then $X^2 - [d]_p$ is irreducible in $\mathbb{Z}_p[X]$, so $\mathbb{Z}_p[X]/(X^2 - [d]_p\,)$ is a field (see problem 3.57)(ii)). Prove that pR is a prime ideal of R and is the only prime ideal P of R with $P \cap \mathbb{Z} = p\mathbb{Z}$.

Note: Dirichlet's Theorem on primes in an arithmetic progression says that for any $a, b \in \mathbb{N}$ with $gcd(a, b) = 1$, there are infinitely many prime numbers p with $p \equiv a \ (mod\ b)$. This is a difficult result in number theory—see, e.g., Janusz [**11**, p. 166] or Borevich & Shafarevich [**3**, pp. 339–341]. (However, the particular case where $a = 1$ has a much easier proof using cyclotomic polynomials—see problem 5.114(iii) below.) From Dirichlet's Theorem it follows easily that there are infinitely many primes in case (c) and infinitely many in case (d). By using Quadratic Reciprocity (see problem 5.128 below), one can calculate rather easily for a given odd prime p whether p is in case (c) or (d) above.

The ring $\mathbb{Z}[\sqrt{-1}\,]$ is called the ring of *Gaussian integers*. For the next three problems, we need that $\mathbb{Z}[\sqrt{-1}\,]$ is a PID. Here is the standard short proof of this: Write $i = \sqrt{-1}$, and recall the norm map $N\colon \mathbb{Z}[i] \to \mathbb{Z}$ of (3.46) given by

$$N(a + bi) \;=\; a^2 + b^2 \;=\; |a + bi|^2.$$

We claim that for any $s, t \in \mathbb{Z}[i]$ with $s \neq 0$,

$$t \;=\; qs + r \quad \text{for some } q, r \in \mathbb{Z}[i] \text{ with } N(r) < N(s). \tag{3.47}$$

For this, write $ts^{-1} = c + di$ with $c, d \in \mathbb{Q}$. (Recall that $\mathbb{Q}[i] = q(\mathbb{Z}[i])$.) Let m be an integer nearest c, and n an integer nearest d; so

$$|c - m| \;\leq\; \tfrac{1}{2} \qquad \text{and} \qquad |d - n| \;\leq\; \tfrac{1}{2}.$$

Let $q = m + ni \in \mathbb{Z}[i]$. Then,

$$|ts^{-1} - q|^2 \;=\; |(c - m) + (d - n)i|^2 \;=\; (c - m)^2 + (d - n)^2 \;\leq\; \tfrac{1}{2}.$$

Hence, for $r = t - qs = (ts^{-1} - q)s \in \mathbb{Z}[i]$,

$$N(r) \;=\; |r|^2 \;=\; |ts^{-1} - q|^2\, |s|^2 \;\leq\; \tfrac{1}{2} N(s) \;<\; N(s),$$

proving (3.47). Now take any nonzero ideal I of $\mathbb{Z}[i]$ and choose $s \in I$ with $N(s)$ minimal among norms of nonzero elements of I. For any $t \in I$, write $t = qs + r$ as in (3.47). Then, $r = t - qs \in I$. Since $N(r) < N(s)$, the choice of s implies that $r = 0$. Thus, $I \subseteq (s)$; since $s \in I$, in fact $I = (s)$. So, every ideal of $\mathbb{Z}[i]$ is principal.

3.61. Let $R = \mathbb{Z}[i] \subseteq \mathbb{C}$, where $i^2 = -1$. Continuing the analysis of the preceding problem, we determine the irreducibles of R. Let q be

any irreducible element of R. Then q is a prime element of the PID R, by problem 3.57, so problem 3.60(vi) applies to the prime ideal qR of R. Let p be the prime number in $\mathbb{N}$ with $qR \cap \mathbb{Z} = p\mathbb{Z}$.

(i) Suppose that $p \neq 2$. Note that $-[1]_p = [p-1]_p \neq [1]_p$. Prove that an element $[a]_p \in \mathbb{Z}_p^*$ satisfies $[a]_p^2 = -[1]_p$ iff $[a]_p$ has order 4 in the group $\mathbb{Z}_p^*$.

(ii) If $p \equiv 3 \ (mod\ 4)$, prove that pR is a prime ideal of R, hence p is irreducible in R. Deduce that $q = up$ for some $u \in R^*$.

(iii) Suppose that $p \equiv 1 \ (mod\ 4)$. Since $\mathbb{Z}_p^*$ is a cyclic group (see problem 3.32), it contains an element of order 4. Deduce that pR is not a prime ideal of R, so p is not a prime element of R, so, as R is a PID, p is not irreducible.

(iv) Suppose that $p \equiv 1 \ (mod\ 4)$ or $p = 2$. Write $p = q_1 q_2 \dots q_k$ where each q_i is irreducible in R and $q_1 = q$. Use the norm map N of (3.46) to prove that $k = 2$ and that $N(q) = p$ and $q_2 = \overline{q}$ (the bar denotes complex conjugate). Prove also that if $q' \in \mathbb{R}$ with $N(q') = p$, then $q' = uq$ or $q' = u\overline{q}$ for some $u \in R^*$. Thus, as $|R^*| = 4$, there are exactly 8 elements q' of R, all irreducible, with $N(q) = p$.

Note that for any prime number p in $\mathbb{N}$, since p is a product of irreducibles in $R = \mathbb{Z}[\sqrt{-1}\,]$, there is an irreducible q of R with $q|p$ in R. Then $qR \cap \mathbb{Z} \supseteq p\mathbb{Z}$; so, as $p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$, we have $qR \cap \mathbb{Z} = p\mathbb{Z}$. Thus, every prime p of $\mathbb{N}$ is covered by parts (ii)–(iv) above. To summarize, the preceding problem shows that the irreducibles of $R = \mathbb{Z}[\sqrt{-1}\,]$ are: (i) up where p is a prime number in $\mathbb{N}$ with $p \equiv 3 \ (mod\ 4)$ and $u \in R^* = \{\pm 1, \pm i\}$; and (ii) elements q of R with $N(q) = p$ for some prime p in $\mathbb{N}$ with $p \equiv 1 \ (mod\ 4)$ or $p = 2$.

3.62. *Sums of two squares in* $\mathbb{N}$. Note that an element of $\mathbb{Z}$ is a sum of two squares of integers iff it is in the image of the norm map $N\colon \mathbb{Z}[\sqrt{-1}\,] \to \mathbb{Z}$ of (3.46). Now take $n \in \mathbb{N}$ with $n \geq 2$, with prime factorization $n = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$ where the p_i are distinct primes and each $r_i \in \mathbb{N}$. Using the preceding problem, prove that n is a sum of two squares of integers iff for each p_i with $p_i \equiv 3 \ (mod\ 4)$ the exponent r_i is even.

3.63. Let $n \in \mathbb{N}$ be a sum of two squares of rational numbers. Prove that n is a sum of two squares of integers.

3.7. Unique factorization domains

An integral domain R is called a *unique factorization domain* (abbreviated UFD) if both

(i) every nonzero nonunit of R is a product of (one or more) irreducible elements;

(ii) every irreducible element of R is a prime element.

Example 3.64. Let R be a PID. Then R is a UFD. Proof: We know from problem 3.57 that every irreducible element of R is a prime element, proving condition (ii) for R. To prove condition (i) by contradiction, suppose there is a nonzero nonunit a of R that is not a product of irreducibles. Since a is not irreducible, it has a nontrivial factorization $a = b_1c_1$ with b_1 and c_1 nonunits and nonzero. Since a is not a product of irreducibles, at least one of b_1 and c_1 cannot be a product of irreducibles; we can assume b_1 is not a product of irreducibles. Note that $(a) \subsetneqq (b_1)$ as $c_1 \notin R^*$. Now iterate this process: Given a nonzero nonunit b_j that is not a product of irreducibles, we can find a factor b_{j+1} of b_j that is not a product of irreducibles with $(b_j) \subsetneqq (b_{j+1})$. Let

$$I = \bigcup_{j=1}^{\infty} (b_j).$$

Because of the inclusion relations among the ideals (b_j), a short calculation shows that I is an ideal of R. Since R is a PID, $I = (d)$ for some $d \in R$. Then, $d \in (b_k)$, for some k. But then, $I = (d) \subseteq (b_k)$, so $b_{k+1} \notin I$ contradicting the definition of I. Hence, there can be no such a, showing that every nonzero nonunit of R is a product of irreducibles. Thus, R is UFD.

3.65. Let R be a UFD, and let a be a nonzero nonunit of R. Then a is a product of irreducibles (which are prime elements of R). Suppose that $a = q_1q_2 \dots q_k$ and $a = q'_1q'_2 \dots q'_\ell$ with the q_i and q'_j all prime elements of R. Prove that $\ell = k$ and that for some permutation $\sigma \in S_k$, we have $q_j \sim q'_{\sigma(j)}$ (i.e., q_j and $q'_{\sigma(j)}$ are associates) for all j.

Thus, the prime factorization of a is unique up to units and the order of the factors. This is why R is called a Unique Factorization Domain.

3.66. Let R be a UFD, and let a be a nonzero nonunit in R, and write $a = q_1 q_2 \dots q_k$ where the q_i are prime elements of R. Take any $b \in R$. Prove that if $b|a$, then $b = u q_{i_1} q_{i_2} \dots q_{i_\ell}$ for some $u \in R^*$ and some subset $\{i_1, i_2, \dots, i_\ell\}$ of $\{1, 2, \dots, k\}$.

Note that the preceding problem allows one to see that greatest common divisors and least common multiples exist for elements of a UFD R, and that they can be read off from the prime factorizations of the elements, just as for $\mathbb{Z}$: Take nonzero $a, b \in R$. We can write

$$a = u q_1^{r_1} q_2^{r_2} \dots q_k^{r_k} \qquad \text{and} \qquad b = v q_1^{s_1} q_2^{s_2} \dots q_k^{s_k},$$

where $u, v \in R^*$ and $q_1, q_2, \dots, q_k$ are pairwise nonassociate prime elements of R, and the r_i and s_i are nonnegative integers. Then,

$$gcd(a, b) \sim q_1^{min(r_1, s_1)} q_2^{min(r_2, s_2)} \dots q_k^{min(r_k, s_k)}.$$

Likewise,

$$lcm(a, b) \sim q_1^{max(r_1, s_1)} q_2^{max(r_2, s_2)} \dots q_k^{max(r_k, s_k)}.$$

3.67. *Rational Roots Test.* Let R be a UFD. Take any

$$f = c_n X^n + c_{n-1} X^{n-1} + \dots + c_1 X + c_0 \in R[X]$$

with $c_0 \neq 0$ and $c_n \neq 0$. Suppose that f has a root s in $q(R)$. We can write $s = a/b$ with $a, b \in R \setminus \{0\}$. Further, by "reducing to lowest terms" (i.e., replacing a and b by $a/gcd(a, b)$ and $b/gcd(a, b)$), we may assume that $gcd(a, b) \sim 1$. Prove that

$$b|c_n \quad \text{and} \quad a|c_0 \quad \text{in } R.$$

(Hint: Clear denominators in the equation $f(a/b) = 0$.) For example, taking $R = \mathbb{Z}$, we can test whether any nonconstant $f \in \mathbb{Z}[X]$ has roots in $\mathbb{Q}$ in finitely many steps, since c_n and c_0 have only finitely many different divisors.

3.68. Let F be any field, and let

$$R = F + X^2 F[X] = \Big\{ \sum_{i=0}^{n} a_i X^i \in F[X] \mid a_1 = 0 \Big\},$$

which is a subring of $F[X]$. Note that $R^* = F[X]^* = F^*$.

(i) Prove that every nonzero non-unit of R is a product of irreducible elements.

(ii) Prove that X^2 is irreducible in R but is not a prime element of R. (Note that X^3 is not a multiple of X^2 in R, as $X \notin R$.) Thus, R is not a UFD.

(iii) Prove that the ideal (X^2, X^3) of R (which equals $X^2F[X]$) is not a principal ideal of R.

(iv) Prove that X^2 and X^3 have no greatest common divisor in R.

(v) Prove that every ideal of R can be generated by two elements.

3.69. Let F be a field, and let $F[X, Y]$ be the polynomial ring in two variables over F. Let $\varepsilon_{F,X^2,X^3}\colon F[X, Y] \to F[X]$ be the evaluation homomorphism sending $f(X, Y)$ to $f(X^2, X^3)$, as in (3.40). Prove that

$$\mathit{im}(\varepsilon_{F,X^2,X^3}) \;=\; F + X^2F[X],$$

the ring of the preceding problem, and that

$$\mathit{ker}(\varepsilon_{F,X^2,X^3}) = (Y^2 - X^3),$$

the principal ideal of $F[X, Y]$ generated by $Y^2 - X^3$. By the FHT,

$$F[X, Y]/(Y^2 - X^3) \;\cong\; F + X^2F[X].$$

Thus, even though $F[X, Y]$ is a UFD by Gauss's Theorem (see problem 3.80 below) its homomorphic image $F[X, Y]/(Y^2 - X^3)$ is an integral domain that is not a UFD by the preceding problem.

3.70. Let

$$R \;=\; \mathbb{Z} + X\mathbb{Q}[X] \;=\; \Big\{ \sum_{i=0}^{n} a_i X^i \in \mathbb{Q}[X] \mid a_0 \in \mathbb{Z} \Big\},$$

which is a subring of $\mathbb{Q}[X]$. Note that since $R^* \subseteq \mathbb{Q}[X]^* = \mathbb{Q}^*$, we have $R^* = \mathbb{Z}^* = \{1, -1\}$.

(i) Prove that the irreducible elements of R are either (a) $\pm p$, where p is a prime number in $\mathbb{N}$; or (b) the $f \in R$ such that $\mathit{deg}(f) \geq 1$, f is irreducible in $\mathbb{Q}[X]$, and $f(0) = \pm 1$.

(ii) Prove that every irreducible element of R is a prime element.

(iii) Prove that X is not expressible as a product of irreducibles in R. Thus, R is not a UFD.

(iv) Prove that every ideal of R generated by two elements is actually a principal ideal. It follows by induction that every ideal of R generated by finitely many elements is a principal ideal.

(v) Deduce that every two elements of R have a greatest common divisor.

(vi) Prove that the ideal $X\mathbb{Q}[X]$ of R cannot be generated by finitely many elements.

3.71. Let R be a UFD. Suppose that for every nonzero a, b in R, any *gcd* d of a and b in R is expressible as $d = ra + sb$ for some $r, s \in R$. Prove that R is a PID.

3.72. The "Fundamental Theorem of Algebra" says that every polynomial in $\mathbb{C}[X]$ of positive degree has a root in $\mathbb{C}$ (or, equivalently, every polynomial in $\mathbb{C}[X]$ of positive degree is a product of polynomials of degree 1). See problem 5.101 below for a proof of this theorem. The problem here gives an equivalent condition in terms of polynomials in $\mathbb{R}[X]$ and is intended to be solved without using the Fundamental Theorem. Prove that the following conditions are equivalent:

(a) Every irreducible polynomial in $\mathbb{C}[X]$ has degree 1.

(b) Every irreducible polynomial in $\mathbb{R}[X]$ has degree 1 or 2.

(Hint: If $f = \sum_{i=0}^{n} c_i X^i \in \mathbb{C}[X]$, let $\overline{f} = \sum_{i=0}^{n} \overline{a_i} X^i$, where $\overline{a_i}$ is the complex conjugate of a_i. Prove that $f\overline{f} \in \mathbb{R}[X]$.)

Localization of an integral domain. Let R be an integral domain, and let S be a nonempty subset of $R \setminus \{0\}$ such that if $s, t \in S$ then $st \in S$. Let

$$R_S = \{rs^{-1} \mid r \in R,\ s \in S\} \subseteq q(R). \tag{3.48}$$

It is easy to check that R_S is a subring of $q(R)$. This R_S is called the *localization of R at S*. Note that if I is an ideal of R, then $I_S = \{is^{-1} \mid i \in I,\ s \in S\}$ is an ideal of R_S. In particular, for $a \in R$, we have $(aR)_S = aR_S$. Also, if J is an ideal of R_S, then $J = (J \cap R)_S$.

Hence, if R is a PID, then R_S is a PID. The next problem shows that if R is a PID, then every subring of $q(R)$ containing R has the form R_S, for some S.

3.73. Let R be a PID, and let T be a subring of $q(R)$ with $T \supseteq R$. Let $S = T^* \cap R$. Observe that S is closed under multiplication and that $0 \notin S$.

(i) Prove that $T = R_S$ as in (3.48). Thus, T is a PID.

(ii) Let $\mathcal{P}(T) = \{p \in R \mid p \text{ is a prime element of } R \text{ and } p \in S\}$. Prove that

$$S = \{\, p_1 p_2 \dots p_n \mid \text{each } p_i \in \mathcal{P}(T),\ n \in \mathbb{N}\} \cup R^*.$$

(iii) Let $\mathcal{P}$ be any nonempty set of prime elements of R, and let

$$\mathcal{S} = \{up_1 p_2 \dots p_n \mid u \in R^*,\ \text{each } p_i \in \mathcal{P},\ n \in \mathbb{N}\},$$

which is a multiplicatively closed subset of R not containing 0. Let $\mathcal{T} = R_{\mathcal{S}}$. Prove that $\mathcal{T}^* \cap R = \mathcal{S}$ and

$$\mathcal{P}(\mathcal{T}) = \{up \mid u \in R^*,\, p \in \mathcal{P}\}.$$

Note that problem 3.73 gives a complete classification of the subrings of $q(R)$ containing R when R is a PID, in terms of the prime elements of R. For example, for every set $\mathcal{P}$ of prime numbers in $\mathbb{N}$, we can build a subring of $\mathbb{Q}$ from $\mathbb{Z}$ and $\mathcal{P}$ as in part (iii). These are all the subrings of $\mathbb{Q}$, and different choices of $\mathcal{P}$ give different rings. In particular, the countable ring $\mathbb{Q}$ has uncountably many different subrings.

3.74. Let $R = \mathbb{Z}[X]$ and let $n \in \mathbb{N}$ with $n \geq 2$. Let T be the subring $\mathbb{Z}[\frac{1}{n}X]$ of $q(R)$. Prove that $T \cong R$, but T is not a localization of R.

3.75. Let R be a UFD, and fix $s \in R \setminus \{0\}$. Let

$$T = R[1/s] = \{r/s^n \mid r \in R \text{ and } n \in \mathbb{N}\},$$

a subring of $q(R)$. (T can also be described as the localization R_S, where $S = \{1, s, s^2, \dots, s^i, \dots\}$.)

(i) Prove that

$$T^* = \{a/s^n \mid a \in R \text{ and } a | s^m \text{ in } R \text{ for some } m, n \in \mathbb{N}\}.$$

(ii) Let $q \in T$. Prove that q is irreducible in T iff $q = up$ for some $u \in T^*$ and p irreducible in R with $p \nmid s$ in R.

(iii) Let $q = up$ and $q' = u'p'$ be irreducibles in T with $u, u' \in T^*$ and p, p' irreducibles of R not dividing s in R. Prove that q and q' are associates in T iff p and p' are associates in R.

(iv) Prove that T is a UFD.

(v) Prove that T is a field iff every irreducible in R divides s.

3.76. *Partial fractions.* One learns in Calculus the method of partial fractions to decompose any quotient of polynomials into a sum of particular kinds of quotients to facilitate integration of the original quotient. This problem gives the corresponding partial fractions decomposition for quotients of polynomials over an arbitrary field F. The quotient field $q(F[X])$ is denoted $F(X)$ and called the *rational function field* over F in the indeterminate X. Recall that the elements of $F(X)$ are formal quotients f/g with $f, g \in F[X]$, with $g \neq 0$, and we have $f/g = f'/g'$ iff $fg' = f'g$. Now fix a particular $r \in F(X)$ with $r \neq 0$.

(i) Prove that there exist unique $f, g \in R[X]$ with g monic and $gcd(f, g) \sim 1$ such that

$$r = f/g.$$

(ii) If $deg(g) = 0$, then $r = f \in F[X]$, and no further decomposition of r is needed. Assume henceforth that $deg(g) \geq 1$. By the Division Algorithm, there are unique $k, h \in F[X]$ with

$$f = kg + h \quad \text{and} \quad deg(h) < deg(g).$$

(Necessarily $h \neq 0$ as $g \nmid f$.) Note that

$$gcd(h, g) \sim gcd(f, g) \sim 1.$$

Then,

$$r = f/g = h/g + k. \tag{3.49}$$

The rest of the process is to decompose h/g as a sum of fractions depending on the prime factorization of g. Suppose first that $g = q^s$ for some $q \in F[X]$ and $s \in \mathbb{N}$. Prove that

there are unique $t_1, t_2, \ldots t_{s-1} \in F[X]$ with $deg(t_i) < deg\, q$ or $t_i = 0$ for each i, such that

$$h = t_1 q^{s-1} + t_2 q^{s-2} + \ldots + t_{s-1} q + t_s.$$

(This can be considered the "base q" representation of h, analogous to the base n representations of positive integers.) Prove also that $t_s \neq 0$. Thus,

$$h/q^s = t_1/q + t_2/q^2 + \ldots + t_{s-1}/q^{s-1} + t_s/q^s.$$

(iii) Now suppose that $g = g_1 g_2$ with $deg(g_i) \geq 1$ for $i = 1, 2$, and $gcd(g_1, g_2) \sim 1$. Thus, $(g_1, g_2) = F[X]$ by problem 3.56. With h as in (3.49), prove that there exist unique h_1 and h_2 in $F[X]$ such that $h = h_1 g_2 + h_2 g_1$ and $deg(h_i) < deg(g_i)$ for $i = 1, 2$. Thus,

$$h/g = h_1/g_1 + h_2/g_2.$$

Prove further that each $h_i \neq 0$ and $gcd(g_i, h_i) \sim 1$.

(iv) Let g have irreducible factorization

$$g = u q_1^{s_1} q_2^{s_2} \ldots q_m^{s_m},$$

where the q_i are pairwise nonassociate irreducibles in $F[X]$, each $s_i \in \mathbb{N}$, and $u \in F[X]^* = F^*$. By multiplying each q_i by a suitable unit in F, we may assume that q_i is monic. Then the q_i are uniquely determined. Since g is monic as well, $u = 1$. Prove that there exist unique $t_{i,j} \in F[X]$ for each $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, s_i\}$ satisfying $deg(t_{i,j}) < deg(q_i)$ or $t_{i,j} = 0$, with $t_{i,s_i} \neq 0$, such that

$$f/g = k + \sum_{i=1}^{m} \sum_{j=1}^{s_i} t_{i,j}/q_i^j,$$

with $k \in F[X]$, as in (3.49). This is the partial fractions decomposition of f/g.

Primitive polynomials. Let R be a UFD. A nonzero polynomial $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ is said to be *primitive* if $gcd(a_0, a_1, \ldots, a_n) \sim 1$. Note that if f is primitive and $b \in R \setminus \{0\}$, then b is a gcd of the coefficients of bf. Also, for any nonzero $g \in R[X]$, if d is a gcd of the coefficients of g, then $g = dg'$ with g' primitive in $R[X]$.

3.77. Let R be a UFD. Prove that for any nonzero $h \in q(R)[X]$, there are $c \in q(R)^*$ and h' primitive in $R[X]$ such that $h = ch'$, and that c is unique up to a multiple in R^*. In particular, prove that if $h \in R[X]$ then $c \in R$.

Gauss's Lemma. Recall one form of *Gauss's Lemma*: If R is a UFD and f and g are primitive polynomials in $R[X]$, then fg is also primitive. (Proof: If fg is not primitive, then there is a prime element q of R dividing all the coefficients of fg. Since f and g are primitive, their images $\overline{f}$ and $\overline{g}$ in $R/(q)[X]$ are nonzero. But, $\overline{f} \cdot \overline{g} = \overline{fg} = 0$. This cannot occur, as $R/(q)[X]$ is an integral domain, since $R/(q)$ is an integral domain.)

3.78. Let $f, g \in \mathbb{Q}[X]$, and suppose that $fg \in \mathbb{Z}[X]$. Prove that the product of any coefficient of f with any coefficient of g lies in $\mathbb{Z}$.

3.79. Let R be a UFD, and let K be its quotient field. Take any nonzero $h \in K[X]$ and write $h = ch'$ where $c \in K^*$ and h' is primitive in $R[X]$. Prove that

$$hK[X] \cap R[X] = h'R[X].$$

3.80. *Gauss's Theorem.* Let R be a UFD, and let K be its quotient field.

(i) Take $f \in R[X]$ with $deg(f) \geq 1$. Prove that f is irreducible in $R[X]$ iff f is primitive in $R[X]$ and f is irreducible in $K[X]$. Prove also that when this occurs f is a prime element of $R[X]$. (Hint: Use the preceding problem.)

(ii) Prove *Gauss's Theorem*: If R is a UFD, then $R[X]$ is also a UFD. Prove also that the irreducible elements of $R[X]$ are the irreducible elements of R together with the primitive polynomials in $R[X]$ of degree ≥ 1 that are irreducible in $K[X]$.

3.81. Let R be a ring. Prove that if $R[X]$ is a UFD, then R is a UFD.

3.82. Let $R = \mathbb{Z}[\sqrt{5/2}\,] \subseteq \mathbb{R}$.

(i) Prove that the evaluation homomorphism $\varepsilon_{\mathbb{Z},\sqrt{5/2}} : \mathbb{Z}[X] \to \mathbb{R}$ has kernel $(2x^2 - 5)\mathbb{Z}[X]$.

(ii) Prove that every element of R is expressible uniquely as $a + b\sqrt{5/2}$, for some $a, b \in \mathbb{Z}[1/2]$.

(iii) Prove that 3 is not a prime element of R, but 7 is a prime element of R.

3.83. Since $\mathbb{Z}[X]$ is a UFD, we know that we can compute gcd's of nonzero elements of $\mathbb{Z}[X]$ from their irreducible factorizations. However, determining the irreducible factorization of $f \in \mathbb{Z}[X]$ (or even determining whether f is irreducible) is generally a difficult computational problem. But determining gcd's in $\mathbb{Z}$ and in $\mathbb{Q}[X]$ is computationally easy because in each case there is a Euclidean Algorithm based on repeated long divisions. This problem shows how one can compute gcd's in $\mathbb{Z}[X]$ by using gcd calculations in $\mathbb{Q}[X]$ and $\mathbb{Z}$.

(i) Let f and g be primitive polynomials in $\mathbb{Z}[X]$, and let h be a gcd of f and g in $\mathbb{Q}[X]$; so h is determined up to a multiple in $\mathbb{Q}^*$. Express $h = ch'$ where $c \in \mathbb{Q}^*$ and h' is primitive in $\mathbb{Z}[X]$. Prove that h' is a gcd of f and g in $\mathbb{Z}[X]$.

(ii) Now take any nonzero $f, g \in \mathbb{Z}[X]$. Write $f = af'$ where a is the gcd of the coefficients of f and f' is primitive in $\mathbb{Z}[X]$. Write $g = bg'$ analogously. Let d be a gcd of a and b in $\mathbb{Z}$, and let h' be a gcd of f' and g' in $\mathbb{Z}[X]$, which is obtainable as in part (i). Prove that dh' is a gcd of f and g in $\mathbb{Z}[X]$.

3.84. *Prime ideals of* $\mathbb{Z}[X]$. Let P be a prime ideal of $\mathbb{Z}[X]$ with $P \neq \{0\}$. If P is also principal ideal of $\mathbb{Z}[X]$, then it is generated by some prime element of $\mathbb{Z}[X]$. From the classification of irreducibles in polynomials over a UFD (see problem 3.80(ii)), we know that then either $P = p\mathbb{Z}[X]$ for some prime number p in $\mathbb{N}$ or $P = f\mathbb{Z}[X]$ for some primitive polynomial f in $\mathbb{Z}[X]$ such that f is irreducible in $\mathbb{Q}[X]$. In this problem we describe all the nonprincipal prime ideals of $\mathbb{Z}[X]$.

(i) Suppose that $P \cap \mathbb{Z} = \{0\}$. Prove that P is a principal ideal. (Hint: Prove that the ideal $P\mathbb{Q}[X]$ of $\mathbb{Q}[X]$ generated by P is a prime ideal of $\mathbb{Q}[X]$, and that $P\mathbb{Q}[X] \cap \mathbb{Z}[X] = P$.)

(ii) Let $p \in \mathbb{N}$ be a prime number, and let $f \in \mathbb{Z}[X]$ be a polynomial whose image $\overline{f}$ in $\mathbb{Z}_p[X]$ is irreducible. Prove that the ideal (p, f) of $\mathbb{Z}[X]$ generated by p and f is a prime ideal

of $\mathbb{Z}[X]$, but is not a principal ideal. Prove also that (p, f) is a maximal ideal of $\mathbb{Z}[X]$.

(iii) Suppose that P is a nonprincipal prime ideal of $\mathbb{Z}[X]$. Since $P \cap \mathbb{Z}$ is a prime ideal of $\mathbb{Z}$, which is not $\{0\}$ by part (i), we must have $P \cap \mathbb{Z} = p\mathbb{Z}$ for some prime number p. Prove that then P has the form (p, f) described in part (ii).

(iv) Prove that no principal prime ideal of $\mathbb{Z}[X]$ is a maximal ideal.

3.85. In the previous problem we saw that every prime ideal of $\mathbb{Z}[X]$ can be generated by at most two elements. The Hilbert Basis Theorem (see, e.g., Dummit & Foote [**5**, p. 316] or Hungerford [**9**, p. 391]) shows that every ideal of $\mathbb{Z}[X]$ can be generated by finitely many elements. This problem shows that for a non-prime ideal there is no upper bound on the number of generators that may be required. Let p be prime number in $\mathbb{N}$, and let $I = (p, X)$ in $\mathbb{Z}[X]$. For $n \in \mathbb{N}$, we have

$$I^n = (p^n, p^{n-1}X, \ldots, p^{n-i}X^i, \ldots, X^n).$$

Prove that I^n has no generating set with fewer than $n + 1$ elements. (Hint: First prove that as an additive group I^n/I^{n+1} is an elementary abelian p-group with $\left|I^n/I^{n+1}\right| = p^{n+1}$. Recall problem 2.53.)

Chapter 4

Linear Algebra and Canonical Forms of Linear Transformations

For the problems in this chapter, it is assumed that the reader is familiar with the most basic linear algebra over the real numbers, including dimension of a vector space, matrix operations, and use of matrices to solve systems of linear equations.

Throughout the chapter, F is any field.

4.1. Vector spaces and linear dependence

Vector spaces. A *vector space* over the field F is an abelian group $(V, +)$ with a *scalar multiplication* operation $\cdot$ of F on V (i.e., a pairing $F \times V \to V$) satisfying, for all $c, d \in F$ and $v, w \in V$,

(i) $c \cdot (v + w) = (c \cdot v) + (c \cdot w)$;

(ii) $(c + d) \cdot v = (c \cdot v) + (d \cdot v)$;

(iii) $(cd) \cdot v = c \cdot (d \cdot v)$;

(iv) $1_F \cdot v = v$.

A vector space over F is also called an F-vector space, or just a vector space when the field F is clear. We write cv for $c \cdot v$, and when parentheses are omitted scalar multiplication takes precedence over addition or subtraction. (Subtraction on V is defined by $v - w = v + -w$.) Thus, $cv + dw - eu$ means $[(c \cdot v) + (d \cdot w)] - (e \cdot u)$.

Let V and W be F-vector spaces. An *F-linear transformation* (or F-vector space homomorphism) from V to W is a function $T\colon V \to W$ such that

$$T(v + v') = T(v) + T(v') \quad \text{and} \quad T(cv) = c(T(v))$$

for all $v, v' \in V$ and $c \in F$. If T is bijective, it is called an *F-vector space isomorphism*. We write $V \cong W$ when there is a vector space isomorphism from V to W.

Example 4.1.

(i) If field F is a subring of a ring R, then R is an F-vector space with the ring multiplication in R used for the scalar multiplication.

(ii) For $m, n \in \mathbb{N}$, let $F^{m\times n}$ denote the set of $m \times n$ matrices over F, with its usual componentwise matrix addition and scalar multiplication. That is, for $A = (a_{ij}) \in F^{m\times n}$, meaning that a_{ij} is the ij-entry of A, and for $c \in F$, define $c \cdot A = (b_{ij})$, where each $b_{ij} = c\,a_{ij}$. Then $F^{m\times n}$ is an F-vector space. We write F^m for $F^{m\times 1}$, called the space of *column vectors* of length m. Also, we write $M_n(F)$ for $F^{n\times n}$.

(iii) Let $V$and W be F-vector spaces. Let $\mathcal{L}_F(V, W)$ be the set of all F-linear transformations from V to W. (When the field F is clear, we write $\mathcal{L}(V, W)$.) For $S, T \in \mathcal{L}(V, W)$ and $c \in F$, define functions $S + T$ and $c \cdot T$ from V to W by

$$(S + T)(v) = S(v) + T(v) \quad \text{and} \quad (c \cdot T)(v) = c \cdot (T(v)).$$

for all $v \in V$. Then $S + T$ and $c \cdot T$ are F-linear transformations. With these operations, $\mathcal{L}(V, W)$ is an F-vector space.

(iv) Let $\{V_i\}_{i\in I}$ be a collection of F-vector spaces. The *direct product* of the V_i is the Cartesian product $\prod_{i\in I} V_i$ with componentwise operations: For any $(\dots, \underset{i}{v_i}, \dots)$, $(\dots, \underset{i}{w_i}, \dots)$ in $\prod_{i\in I} V_i$ and $c \in F$, set

$$(\dots, \underset{i}{v_i}, \dots) + (\dots, \underset{i}{w_i}, \dots) = (\dots, \underset{i}{v_i + w_i}, \dots) \tag{4.1}$$

$$\text{and} \quad c \cdot (\dots, \underset{i}{v_i}, \dots) = (\dots, \underset{i}{c\, v_i}, \dots) \tag{4.2}$$

With these operations, $\prod_{i\in I} V_i$ is an F-vector space.

(v) Let $\{V_i\}_{i\in I}$ be a collection of F-vector spaces. The *direct sum* of the V_i is the subset of the direct product,

$$\bigoplus_{i\in I} V_i = \left\{ (\dots, \underset{i}{v_i}, \dots) \in \prod_{i\in I} V_i \;\middle|\; \begin{array}{l}\text{at most finitely many} \\ v_i \text{ are nonzero}\end{array} \right\}. \tag{4.3}$$

With the operations as in (4.1), $\bigoplus_{i\in I} V_i$ is an F-vector space.

4.2. Let W be an abelian group. Recall the endomorphism ring $End(W)$ of group homomorphisms $W \to W$, see (3.2).

(i) Suppose there is a ring homomorphism $\beta\colon F \to End(W)$. Define a scalar multiplication of F on W by

$$c \cdot w = \beta(c)(w)$$

for all $c \in F$, $w \in W$. Prove that with this scalar multiplication and addition given by the group operation on W, W is an F-vector space.

(ii) Conversely, suppose that the group W is an F-vector space. Define a function $\alpha\colon F \to End(W)$ by

$$\alpha(c)(w) = c \cdot w$$

for all $c \in F$, $w \in W$. Prove that α is a ring homomorphism.

Subspaces and factor spaces. Let V be an F-vector space. A nonempty subset W of V is called an *F-subspace of V* if $(W, +)$ is a subgroup of $(V, +)$ and $cw \in W$ for all $c \in F$, $w \in W$. When F is clear, W is called simply a subspace. The set $\{0_V\}$ is the *trivial*

subspace of V. Any intersection of subspaces of V is again subspace of V. If $W_1, W_2, \ldots, W_n$ are subspaces of V, then

$$W_1 + W_2 + \ldots + W_n \;=\; \{w_1 + w_2 + \ldots + w_n \mid \text{each } w_i \in W_i\} \quad (4.4)$$

is the subspace of V generated by $W_1, \ldots, W_n$.

Let W be any subspace of the F-vector space V. We have the factor group of the additive group

$$V/W \;=\; \{v{+}W \mid v \in V\}, \quad \text{where} \quad v{+}W = \{v{+}w \mid w \in W\}. \quad (4.5)$$

Recall that

$$v_1 + W \;=\; v_2 + W \quad \text{iff} \quad v_1 - v_2 \;\in\; W. \quad (4.6)$$

There are well-defined operations of addition and scalar multiplication on V/W given by

$$(v + W) + (v' + W) = (v + v') + W \;\text{ and }\; c \cdot (v + W) = (cv) + W$$

for all $c \in F$, $v, v' \in V$, making V/W into an F-vector space; V/W is called the *factor space* of V modulo W.

4.3. *Elementary abelian p-groups as $\mathbb{Z}_p$-vector spaces.* Let p be a prime number. Recall that $\mathbb{Z}_p$ is a field. Let $(A, +)$ be an abelian group, and suppose that $pa = 0_A$ for every $a \in A$.

(i) Prove that the scalar multiplication of $\mathbb{Z}_p$ on A given by

$$[i]_p \cdot a \;=\; ia \qquad \text{for all } i \in \mathbb{Z},\ a \in A$$

is well-defined and makes A into a $\mathbb{Z}_p$-vector space. (For a quick proof of this, use problem 4.2.)

(ii) Prove that every subgroup of A is a $\mathbb{Z}_p$-subpace.

(iii) Suppose that $(B, +)$ is another abelian group with $pb = 0$ for every $b \in B$. Prove that every group homomorphism from A to B is a $\mathbb{Z}_p$-linear transformation; thus,

$$\mathcal{L}_{\mathbb{Z}_p}(A, B) \;=\; \mathit{Hom}(A, B).$$

Note that the finite abelian groups $(A, +)$ satisfying $pa = 0_A$ for all $a \in A$ are the elementary abelian p-groups as in problem 2.53.

Let V be an F-vector space, and take any $v_1, \ldots, v_n \in V$. A *linear combination* of the v_i is any sum $c_1v_1 + c_2v_2 + \ldots + c_nv_n$,

where $c_1, \ldots, c_n \in F$. The *span* of the v_i is the set of all such linear combinations,

$$span\{v_1, \ldots, v_n\} \;=\; \{c_1v_1 + c_2v_2 + \ldots + c_nv_n \mid c_1, \ldots, c_n \in F\}, \quad (4.7)$$

which is the F-subspace of V generated by the v_i. We say that $v_1, \ldots, v_n$ are *linearly independent* when $c_1v_1 + c_2v_2 + \ldots + c_nv_n = 0$ implies that $c_1 = c_2 = \ldots = c_n = 0$. This is equivalent to: whenever

$$c_1v_1 + c_2v_2 + \ldots + c_nv_n \;=\; d_1v_1 + d_2v_2 + \ldots + d_nv_n$$

with $c_1, \ldots, c_n, d_1, \ldots, d_n \in F$, we have $c_1 = d_1$, $c_2 = d_2$, $\ldots$, $c_n = d_n$. Note that $v_1, \ldots, v_n$ are linearly independent iff $v_1 \neq 0$ and $v_i \notin span\{v_1, v_2, \ldots, v_{i-1}\}$ for $i \in \{2, 3, \ldots, n\}$. If $\{v_i\}_{i\in I}$ is an infinite subset of V, a linear combination of the v_i is a linear combination of some finite subset of the set of v_i. The span of the v_i is the sum (which is actually the union) of the spans of finite subsets of the set of v_i. The v_i are linearly independent if every finite subset of $\{v_i\}_{i\in I}$ is linearly independent. We say that V is *finitely-generated* if it is the span of some finite subset of V.

A subset $\{v_i\}_{i\in I}$ of the vector space V is a *base* (or basis) of V if the subset is linearly independent and also spans V. Note that if $\{w_j\}_{j\in J}$ spans V, then any maximal linearly independent subset of the $\{w_j\}_{j\in J}$ is a base of V. This shows that any finitely-generated vector space has a finite base. Recall the basic theorem that if V is finitely-generated, then any two bases of V have the same (finite) number of elements. This number is called the *dimension of* V, denoted $dim_F(V)$ (or $dim(V)$ when the field F is clear). If $V = \{0_V\}$, then $\varnothing$ is considered a base of V, and $dim(\{0_V\})) = 0$. If V is nontrivial and finitely-generated, then $dim(V) \in \mathbb{N}$. Note that $dim(V)$ is an upper bound on the cardinality of any linearly independent subset of V, since such a subset can be enlarged to yield a base of V. Also, $dim(V)$ is a lower bound on the cardinality of any subset of V spanning V, since any spanning set contains a subset which is a base of V.

If V is not finitely-generated, then Zorn's Lemma shows that V contains a maximal linearly independent subset (see §0.2 above); such a subset is clearly a base of V, necessarily of infinite cardinality. We then write $dim(V) = \infty$.

Example 4.4. Taken any $m, n \in \mathbb{N}$. In $F^{m\times n}$, let E_{ij} be the matrix with ij-entry 1 and all other entries 0. For any $A = (a_{ij}) \in F^{m\times n}$, we have

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} E_{ij},$$

and this is the unique way of expressing A as a linear combination of the E_{ij}. Thus, $\{E_{ij}\}_{i=1,j=1}^{m\ n}$ is a base of $F^{m\times n}$; so, $dim(F^{m\times n}) = mn$.

Abstract dependence relations. Let S be a set. A relation $\prec$ between elements of S and subsets of S is called a *dependence relation* if it satisfies the following axioms, for all $s \in S$ and subsets T, U of S:

(i) If $s \in T$, then $s \prec T$.

(ii) If $s \prec T$ and every $t \in T$ satisfies $t \prec U$, then $s \prec U$.

(iii) If $s \prec T$, then there is some finite subset T_0 of T with $s \prec T_0$.

(iv) If $s \prec T$ but $s \not\prec T\backslash\{t\}$, for some $t \in T$, then $t \prec (T\backslash\{t\}) \cup \{s\}$.

When $s \prec T$ we say that s is *dependent on* T. A subset I of S is *independent* if for each $s \in I$ we have $s \not\prec I \setminus \{s\}$. For example, trivially $\varnothing$ is independent. A subset U of S *spans* S if $s \prec U$ for each $s \in S$. For example, S spans S. A subset B of S is a *base* of S if B is independent and spans S.

4.5. Let S be a set with a dependence relation $\prec$.

(i) Let T and Y be subsets of S with $T \subseteq Y$. For $s \in S$ prove that if $s \prec T$ then $s \prec Y$. Deduce that if Y is independent, then T is independent.

(ii) Let I be an independent subset of S and take any $s \in S$ with $s \not\prec I$. Prove that $I \cup \{s\}$ is independent.

(iii) Prove that a subset I of S is independent iff every finite subset of I is independent.

(iv) Let I be an independent subset of S, and let U be a subset of S such that U spans S and $I \subseteq U$. Prove (using Zorn's Lemma if U is infinite) that there is a set B with $I \subseteq B \subseteq U$ such that B is maximal among the independent subsets of U that contain I. Prove that any such B is a base of S.

(v) Let B be a finite base of S, and let U be a subset of S such that U spans S. Prove that there is a base B' of S

with $B' \subseteq U$ and $|B'| = |B|$. (Hint: Argue by induction on $|B \setminus (B \cap U)|$: If $B \not\subseteq U$, take any $b \in B \setminus (B \cap U)$ and show that there is $u \in U$ with $u \not\prec B \setminus \{b\}$, and that $(B \setminus \{b\}) \cup \{u\}$ is a base of S.)

(vi) Let B be a base of S with $|B| < \infty$. Apply part (v) to prove that if B' is any other base of S, then $|B'| = |B|$. (It follows from this and previous parts that if I is any independent set then $|I| \leq |B|$ and if U is any subset of S that spans S then $|B| \leq |U|$.)

Note also that if B is a base of S with $|B| = \infty$ and B' is any other base of S, then $|B'| = |B|$ as infinite cardinal numbers. For, for each $b' \in B'$ there is a finite subset $C_{b'}$ of B such that $b' \prec C_{b'}$. Let $D = \bigcup_{b' \in B'} C_{b'} \subseteq B$. Then, D spans S by axiom (ii), as B' spans S and $b' \prec D$ for each $b' \in B'$. Since in addition $D \subseteq B$ with B independent, we must have $D = B$. Thus,

$$|B| = |D| \geq \aleph_0 |B'|.$$

Moreover, if B' were finite then D would be finite, contrary to the assumption on $|B|$; so $|B'| = \infty$. Likewise, $|B'| \geq \aleph_0 |B|$; hence $|B'| = |B|$.

4.6. Let V be a vector space. Define a relation $\prec$ between elements v of V and subsets T of V by: $v \prec T$ just when v is a linear combination of elements of T. Prove that $\prec$ is a dependence relation. Note that for this dependence relation the independent subsets of V are the linearly independent subsets of V; spanning sets are subsets that span V as a vector space; and bases are vector space bases of V. Thus, problem 4.5 proves the existence of vector space bases of V, and that any two bases have the same cardinality.

4.7. Let V be a finite-dimensional F-vector space, and let W be a subspace of V. Suppose that $\{w_1, w_2, \ldots, w_m\}$ is a base of W. Take any $v_1, \ldots, v_n \in V$. Write $\overline{v_i}$ for the image $v_i + W$ of v_i in V/W.

(i) Prove that $\overline{v_1}, \overline{v_2}, \ldots, \overline{v_n}$ are linearly independent in V/W iff $v_1, \ldots, v_n, w_1, \ldots, w_m$ are linearly independent in V.

(ii) Prove that $\overline{v_1}, \overline{v_2}, \ldots, \overline{v_n}$ span V/W iff $v_1, \ldots, v_n, w_1, \ldots, w_m$ span V. It follows from this and part (i) that $\{\overline{v_1}, \ldots, \overline{v_n}\}$

is a base of V/W iff $\{v_1, \ldots, v_n, w_1, \ldots, w_m\}$ is a base of V. Hence,

$$dim(W) + dim(V/W) \;=\; dim(V). \tag{4.8}$$

4.2. Linear transformations and matrices

Let V and W be F-vector spaces, and let $T\colon V \to W$ be an F-linear transformation. The *image* of T (also called the *range* of T) is

$$im(T) \;=\; \{T(v) \mid v \in V\}.$$

Note that $im(T)$ is a vector subspace of W. The dimension of $im(T)$ is called the *rank* of T, and denoted $rk(T)$. The *kernel* of T (also called the *nullspace* of T) is

$$ker(T) \;=\; \{v \in V \mid T(v) = 0_W\}, \tag{4.9}$$

which is the same as the kernel of T as an additive group homomorphism $(V,+) \to (W,+)$. Note that $ker(T)$ is a vector subspace of V. The dimension of $ker(T)$ is sometimes called the *nullity* of T. There is a Fundamental Homomorphism Theorem for linear transformations of vector spaces, analogous to the FHT for groups and for rings, and deducible easily from the FHT for groups. It yields in particular that for any linear transformation $T\colon V \to W$, the additive group isomorphism

$$V/\,ker(T) \;\cong\; im(T), \tag{4.10}$$

is a vector space isomorphism. When V is finite-dimensional, the FHT combined with (4.8) above yield the *Dimension Theorem*

$$dim(V) \;=\; dim(ker(T)) + rk(T). \tag{4.11}$$

The FHT for vector spaces yields vector space analogues to the First and Second Isomorphism Theorems and the Correspondence Theorem for groups and rings. In particular, the First Isomorphism Theorem says: For vector subspaces W_1 and W_2 of a vector space V,

$$W_1/(W_1 \cap W_2) \;\cong\; (W_1 + W_2)/W_2. \tag{4.12}$$

Hence, when W_1 and W_2 are finite-dimensional, (4.8) yields

$$dim(W_1 + W_2) + dim(W_1 \cap W_2) \;=\; dim(W_1) + dim(W_2). \tag{4.13}$$

Internal direct sums. Let V be an F-vector space, and let $W_1, W_2, \ldots, W_n$ be subspaces of V, with $n \geq 2$. There is a natural linear transformation $T\colon W_1 \oplus W_2 \oplus \ldots \oplus W_n \to V$ given by

$$T\big((w_1, w_2, \ldots, w_n)\big) \;=\; w_1 + w_2 + \ldots + w_n.$$

When T is an isomorphism, we say that V is the (internal) *direct sum* of the W_i, and write

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_n. \tag{4.14}$$

4.8. Let V be an F-vector space, and let $W_1, W_2, \ldots, W_n$ be subspaces of V, with $n \geq 2$. Suppose that $W_1 + W_2 + \ldots + W_n = V$. Prove that the following conditions are equivalent.

(a) $V = W_1 \oplus W_2 \oplus \ldots \oplus W_n$.

(b) Every $v \in V$ is expressible uniquely as $v = w_1 + w_2 + \ldots + w_n$ with each $w_i \in W_i$.

(c) If $\mathcal{B}_i$ is a base of W_i for each i, then $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_n$ is a base of V.

(d) $W_i \cap (W_1 + \ldots + W_{i-1}) = \{0\}$ for $i = 2, 3, \ldots, n$.

(e) $W_j \cap (W_1 + \ldots + W_{j-1} + W_{j+1} + \ldots + W_n) = \{0\}$ for $j = 1, 2, \ldots, n$.

If V is finite-dimensional, prove that (a)–(e) are equivalent to

(f) $\dim(V) = \dim(W_1) + \dim(W_2) + \ldots + \dim(W_n)$.

Let V and W be F-vector spaces. Recall from Example 4.1(iii) the vector space $\mathcal{L}(V, W)$ of F-linear transformations from V to W. Note that if $\mathcal{B} = \{v_i\}_{i \in I}$ is a base of V and $\{w_i\}_{i \in I}$ is any subset of W, then there is a unique $T \in \mathcal{L}(V, W)$ with $T(v_i) = w_i$ for each $i \in I$. When the two vector spaces are the same, we write $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$. Note that $\mathcal{L}(V)$ is a ring as well as a vector space, with addition as above and multiplication given by $S \cdot T = S \circ T$. Also, $1_{\mathcal{L}(V)} = id_V$, the identity map on V.

4.9. Let V and W be F-vector spaces, and let $T \in \mathcal{L}(V, W)$.

(i) An $S \in \mathcal{L}(W, V)$ is a *left inverse* of T if $S \circ T = id_V$. Prove that T has a left inverse iff $\ker(T) = \{0_V\}$.

(ii) A $U \in \mathcal{L}(W,V)$ is a *right inverse* of T if $T \circ U = id_W$. Prove that T has a right inverse iff $im(T) = W$.

(iii) T is said to be *invertible* if it has both a left inverse and a right inverse. Prove that when this occurs, the left inverse is unique, as is the right inverse, and they coincide. This map is called the inverse of T, and denoted T^{-1}.

It follows from parts (i) and (ii) and the Dimension Theorem (4.11) that if $dim(V) = dim(W) < \infty$, then T is invertible iff it has a left inverse, iff it has a right inverse.

4.10. Let V be an F-vector space with a countably infinite base (e.g., $V = F[X]$).

(i) Give an example of $T \in \mathcal{L}(V)$ with a left inverse but no right inverse.

(ii) Give an example of $S \in \mathcal{L}(V)$ with a right inverse but no left inverse.

(iii) Let $I = \{T \in \mathcal{L}(V) \mid rk(T) < \infty\}$. Prove that I is an ideal of $\mathcal{L}(V)$, and that it is the only ideal besides $\{0\}$ and $\mathcal{L}(V)$.

4.11. *Lagrange interpolation.* Take any distinct elements $a_1, a_2, \ldots, a_n$ in the field F. Let V be the n-dimensional F-subspace of $F[X]$ with base $\{1, X, X^2, \ldots X^{n-1}\}$, i.e., V is the set of polynomials of degree at most $n-1$, together with 0. Define a linear transformation $T\colon V \to F^{1\times n}$ by

$$T(f) = \big(f(a_1), f(a_2), \ldots, f(a_n)\big).$$

Prove that T is a vector space isomorphism. Thus, for any $b_1, b_2, \ldots, b_n \in F$, there is a unique $f \in F[X]$ of degree $\leq n-1$ (or $f = 0$) such that $f(a_i) = b_i$ for each i. There is actually an explicit formula for f, as follows:

$$\begin{aligned} f &= \sum_{i=1}^{n} \tfrac{b_i(X-a_1)(X-a_2)\ldots(X-a_{i-1})(X-a_{i+1})\ldots(X-a_n)}{(a_i-a_1)(a_i-a_2)\ldots(a_i-a_{i-1})(a_i-a_{i+1})\ldots(a_i-a_n)} \\ &= \sum_{i=1}^{n} \big(b_i \prod_{j\neq i} \tfrac{X-a_j}{a_i-a_j}\big). \end{aligned} \tag{4.15}$$

This expression is known as *Lagrange's interpolation formula.* It is intended that you solve this problem without using Lagrange's formula.

4.12. Let V, W, and Y be three finite-dimensional F-vector spaces. Let $T \in \mathcal{L}(V,W)$ and $S \in \mathcal{L}(W,Y)$, so $S \circ T \in \mathcal{L}(V,Y)$.

(i) Prove that

$$\begin{aligned} rk(S \circ T) &= rk(T) - dim\big(\, im(T) \cap ker(S)\big) \\ &= dim\big(\, im(T) + ker(S)\big) - dim(ker(S)). \end{aligned}$$

(ii) Deduce that

$$\begin{aligned} rk(T) + rk(S) - dim(W) &\leq rk(S \circ T) \\ &\leq min\big(rk(T), rk(S)\big). \end{aligned} \tag{4.16}$$

Invertible matrices. Recall that a square matrix $A \in M_n(F)$ is said to be *invertible* if there is a matrix $B \in M_n(F)$ with

$$BA = AB = I_n,$$

where I_n is the identity matrix in $M_n(F)$. When such a B exists, it is uniquely determined, and we write $B = A^{-1}$.

Coordinate vectors and change of base matrices. Let V be a finite-dimensional F-vector space, and let $n = dim(V)$. Fix a base $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ of V. Any $v \in V$ is expressible uniquely as

$$v = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n,$$

with the $c_i \in F$. The *coordinate vector of* v relative to the base $\mathcal{B}$ is

$$[v]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in F^n. \tag{4.17}$$

In this context, $\mathcal{B}$ is understood to be an *ordered base* of V, i.e., a base with a specified ordering of the vectors in the base. Note that the map $V \to F^n$ given by $v \mapsto [v]_{\mathcal{B}}$ is a vector space isomorphism. Let $\mathcal{C} = \{w_1, w_2, \ldots, w_n\}$ be another (ordered) base of V. The *change of base matrix* from $\mathcal{B}$ to $\mathcal{C}$ is the matrix $P \in M_n(F)$ with j-th column $[w_j]_{\mathcal{B}}$ for $j = 1, 2, \ldots, n$. Note that for any $v \in V$,

$$[v]_{\mathcal{B}} = P[v]_{\mathcal{C}}. \tag{4.18}$$

If $\mathcal{D}$ is another base of V, and S is the base change matrix from $\mathcal{C}$ to $\mathcal{D}$, then $[v]_{\mathcal{B}} = PS[v]_{\mathcal{D}}$ by (4.18), so PS is the base change matrix from $\mathcal{B}$ to $\mathcal{D}$. In particular, if Q is the base change matrix from $\mathcal{C}$ to $\mathcal{B}$, then PQ is the base change matrix from $\mathcal{B}$ to $\mathcal{B}$, so $PQ = I_n$, the

identity matrix in $M_n(F)$; likewise, $QP = I_n$. Hence, P is invertible in $M_n(F)$, and $Q = P^{-1}$.

Matrix of a linear transformation. Let V and W be finite-dimensional F-vector spaces, and let $T \in \mathcal{L}(V, W)$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an (ordered) base of V, and $\mathcal{C} = \{w_1, \ldots, w_m\}$ an (ordered) base of W. The *matrix of* T relative to the bases $\mathcal{B}$ and $\mathcal{C}$ is

$$[T]_{\mathcal{B}}^{\mathcal{C}} = (a_{ij}) \in F^{m \times n} \quad \text{where} \quad T(v_j) = \sum_{i=1}^{m} a_{ij} w_i, \tag{4.19}$$

for $j = 1, 2, \ldots, n$. In terms of column vectors, the j-th column of $[T]_{\mathcal{B}}^{\mathcal{C}}$ is $[T(v_j)]_{\mathcal{C}}$. Furthermore, for any $v \in V$,

$$[T(v)]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}} [v]_{\mathcal{B}}. \tag{4.20}$$

For fixed bases $\mathcal{B}$ and $\mathcal{C}$, the map $T \mapsto [T]_{\mathcal{B}}^{\mathcal{C}}$ yields a vector space isomorphism

$$\mathcal{L}(V, W) \cong F^{m \times n}.$$

Hence, $dim(\mathcal{L}(V, W)) = dim\, F^{m \times n} = mn$.

Still assuming that V and W are finite-dimensional F-vector spaces, let $\mathcal{B}$ and $\mathcal{B}'$ be two bases of V, and let $\mathcal{C}$ and $\mathcal{C}'$ be two bases of W. Let P be the change of base matrix from $\mathcal{B}$ to $\mathcal{B}'$ and Q the change of base matrix from $\mathcal{C}$ to $\mathcal{C}'$ as in (4.18). Take any $T \in \mathcal{L}(V, W)$. Then, for any $v \in V$, we have

$$\begin{aligned} Q^{-1}[T]_{\mathcal{B}}^{\mathcal{C}} P\, [v]_{\mathcal{B}'} &= Q^{-1}[T]_{\mathcal{B}}^{\mathcal{C}} [v]_{\mathcal{B}} = Q^{-1}[T(v)]_{\mathcal{C}} \\ &= [T(v)]_{\mathcal{C}'} = [T]_{\mathcal{B}'}^{\mathcal{C}'} [v]_{\mathcal{B}'}. \end{aligned}$$

This yields the base change formula

$$[T]_{\mathcal{B}'}^{\mathcal{C}'} = Q^{-1}[T]_{\mathcal{B}}^{\mathcal{C}} P. \tag{4.21}$$

Note also that if U is another finite-dimensional vector space, and $S \in \mathcal{L}(W, U)$ and $\mathcal{D}$ is a base of U, then

$$[ST]_{\mathcal{B}}^{\mathcal{D}} = [S]_{\mathcal{C}}^{\mathcal{D}} [T]_{\mathcal{B}}^{\mathcal{C}}. \tag{4.22}$$

If $W = V$, then we typically use the same base $\mathcal{B}$ of V for both the domain of $T \in \mathcal{L}(V)$ and the target of T, and write

$$[T]_{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}}. \tag{4.23}$$

Thus, if $\mathcal{B}'$ is another base of V and P is the base change matrix from $\mathcal{B}$ to $\mathcal{B}'$, then by (4.21)

$$[T]_{\mathcal{B}'} \;=\; P^{-1}[T]_{\mathcal{B}}\, P. \tag{4.24}$$

Similar matrices. Two matrices $A, B \in M_n(F)$ are said to be *similar* if there is an invertible matrix $P \in M_n(F)$ such that

$$B \;=\; P^{-1}AP.$$

It is easy to see that similarity is an equivalence relation on $M_n(F)$. Equation (4.24) shows that any two matrices representing a linear transformation $T \in \mathcal{L}(V)$ are similar.

Matrices as linear transformations. For any $m, n \in \mathbb{N}$, take any $A \in F^{m\times n}$. Then, left multiplication by A gives a linear transformation $L_A \in \mathcal{L}(F^n, F^m)$. Thus, for $v \in F^n$,

$$L_A(v) \;=\; Av \;\in\; F^m.$$

Let $\mathcal{S} = \{\varepsilon_1, \ldots, \varepsilon_n\}$ be the standard base of F^n, i.e, $\varepsilon_i \in F^n$ is the column vector with i-entry 1 and all other entries 0; let $\mathcal{S}'$ be the analogous base of F^m. Note that

$$[L_A]_{\mathcal{S}}^{\mathcal{S}'} \;=\; A. \tag{4.25}$$

Column space. For any $m \in \mathbb{N}$, take any column vectors $\beta_1, \beta_2, \ldots, \beta_n \in F^m$. Set

$$[\beta_1, \beta_2, \ldots, \beta_n] \;=\; \text{the matrix in } F^{m\times n} \text{ with } j\text{-th column } \beta_j. \tag{4.26}$$

Let $B = [\beta_1, \beta_2, \ldots, \beta_n]$. The *column space* of B is

$$\mathit{Col}(B) \;=\; \mathit{span}\{\beta_1, \ldots, \beta_n\} \;\subseteq\; F^m. \tag{4.27}$$

The *column rank* of B is

$$\mathit{col\text{-}rk}(B) \;=\; \mathit{dim}(\mathit{Col}(B)). \tag{4.28}$$

The *row space* of B, $\mathit{Row}(B) \subseteq F^{1\times n}$, and its *row rank*, $\mathit{row\text{-}rk}(B)$, are defined analogously. Let $A \in F^{s\times m}$. Then for $B = [\beta_1, \beta_2, \ldots, \beta_n]$, we have

$$AB \;=\; [A\beta_1, A\beta_2, \ldots, A\beta_n].$$

Since each $A\beta_j$ is a linear combination of the columns of A, we have

$$\mathit{Col}(AB) \;\subseteq\; \mathit{Col}(A), \quad \text{hence} \quad \mathit{col\text{-}rk}(AB) \;\leq\; \mathit{col\text{-}rk}(A),$$

with equality if B has a right inverse. Also, if $\beta_{j_1}, \ldots, \beta_{j_k}$ span $Col(B)$, then $A\beta_{j_1}, \ldots, A\beta_{j_k}$ span $Col(AB)$; hence,

$$col\text{-}rk(AB) \leq col\text{-}rk(B),$$

with equality if A has a left inverse. Analogous formulas hold for $row\text{-}rk(AB)$. Note that if $T \in \mathcal{L}(V, W)$ for finite-dimensional vector spaces V and W with respective bases $\mathcal{B}$ and $\mathcal{C}$, then from (4.20) we have $\{[T(v)]_{\mathcal{C}} \mid v \in V\} = Col([T]^{\mathcal{C}}_{\mathcal{B}})$; hence,

$$col\text{-}rk([T]^{\mathcal{C}}_{\mathcal{B}}) = rk(T). \tag{4.29}$$

4.13. Take any $T \in \mathcal{L}(V, W)$ for any finite-dimensional F-vector spaces V and W, with $dim(V) = n$ and $dim(W) = m$. Let $r = rk(T)$. Prove that there are bases $\mathcal{B}$ of V and $\mathcal{C}$ of W such that $[T]^{\mathcal{C}}_{\mathcal{B}} \in F^{m\times n}$ has block form

$$[T]^{\mathcal{C}}_{\mathcal{B}} = \begin{pmatrix} I_r & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_3 \end{pmatrix}, \tag{4.30}$$

where I_r is the identity matrix in $M_r(F)$, and $\mathbf{0}_1$, $\mathbf{0}_2$, and $\mathbf{0}_3$ are the 0-matrices in $F^{r\times(n-r)}$, $F^{(m-r)\times r}$, and $F^{(m-r)\times(n-r)}$.

4.14. Take any $A \in F^{m\times n}$ for any $m, n \in N$. Prove the *rank equality*:

$$row\text{-}rk(A) = col\text{-}rk(A). \tag{4.31}$$

(Hint: Apply the preceding problem to $L_A \in \mathcal{L}(F^n, F^m)$.)

See problem 4.20 below for another approach to proving the rank equality (4.31).

4.15. Let V be a finite-dimensional F-vector space. Take any ring and vector space isomorphism $\psi\colon \mathcal{L}(V) \to \mathcal{L}(V)$. Prove that there is an invertible $S \in \mathcal{L}(V)$ such that $\psi(T) = S^{-1}TS$ for all $T \in \mathcal{L}(V)$.

Invariant subspaces. Let V be a finite-dimensional F-vector space, and let $T \in \mathcal{L}(V)$. A subspace W of V is said to be T-*invariant* if $T(w) \in W$ for every $w \in W$. When this occurs, there are well-defined associated linear transformations $T|_W \in \mathcal{L}(W)$ (the restriction of T to W) and $\overline{T} \in \mathcal{L}(V/W)$ defined by

$$T|_W(w) = T(w), \quad \text{for all } w \in W$$

and

$$\overline{T}(\overline{v}) = \overline{T(v)} \quad \text{for all } \overline{v} = v + W \in V/W.$$

Let $\mathcal{L}_W(V)$ be the set of all linear transformations $T \in \mathcal{L}(V)$ such that W is T-invariant. Observe that $\mathcal{L}_W(V)$ is a subring and vector subspace of $\mathcal{L}(V)$, and that the maps $\mathcal{L}_W(V) \to \mathcal{L}(W)$ given by $T \mapsto T|_W$ and $\mathcal{L}_W(V) \to \mathcal{L}(V/W)$ given by $T \mapsto \overline{T}$ are ring and F-vector space homomorphisms. Suppose $dim(V) = n$ and $dim(W) = r$. Take any base $\mathcal{C} = \{w_1, \ldots, w_r\}$ of W, and enlarge it to a base $\mathcal{B} = \{w_1, \ldots, w_r, v_1, \ldots, v_{n-r}\}$ of V. Recall from problem 4.7 that $\overline{\mathcal{D}} = \{\overline{v_1}, \ldots, \overline{v_{n-r}}\}$ is a base of V/W. Note that for such a base $\mathcal{B}$, the matrix $[T]_{\mathcal{B}}$ in $M_n(F)$ has block triangular form

$$[T]_{\mathcal{B}} \;=\; \begin{pmatrix} [T|_W]_{\mathcal{C}} & B \\ \mathbf{0} & [\overline{T}]_{\overline{\mathcal{D}}} \end{pmatrix}, \tag{4.32}$$

for some $B \in F^{r\times(n-r)}$, where $\mathbf{0}$ denotes the 0-matrix in $F^{(n-r)\times r}$.

For $A \in M_n(F)$, an *A-invariant subspace* of F^n is an invariant subspace for the left multiplication by A map $L_A \in \mathcal{L}(F^n)$.

4.3. Dual space

Throughout this section, let V be a finite-dimensional F-vector space.

Transpose matrix. Let $A = (a_{ij}) \in F^{m\times n}$. The *transpose* of A is the matrix $A^t \in F^{n\times m}$ such that the ij-entry of A^t is a_{ji}, for all i, j. Note that if $B \in F^{m\times n}$ and $C \in F^{s\times m}$, then

$$(A+B)^t \;=\; A^t + B^t \qquad \text{and} \qquad (CA)^t \;=\; A^t C^t.$$

Dual spaces of vector spaces. The *dual space* of the finite-dimensional vector space V is

$$V^* \;=\; \mathcal{L}(V, F). \tag{4.33}$$

Note that if $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ is any base of F, then there is a corresponding *dual base* $\mathcal{B}^*$ of V^* given by

$$\mathcal{B}^* \;=\; \{v_1^*, v_2^*, \ldots, v_n^*\}, \quad \text{where} \quad v_i^*(v_j) \;=\; \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{if } j \neq i. \end{cases} \tag{4.34}$$

Thus,

$$dim(V^*) \;=\; dim(V). \tag{4.35}$$

If W is another finite-dimensional F-vector space and $T \in \mathcal{L}(V,W)$, then there is a *dual linear transformation* $T^* \in \mathcal{L}(W^*,V^*)$ defined by

$$T^*(y)(v) \;=\; y(T(v)) \qquad \text{for all } \; y \in W^*, v \in V. \tag{4.36}$$

Note that if $\mathcal{B}$ is a base of V and $\mathcal{C}$ is a base of W, then

$$[T^*]^{\mathcal{B}^*}_{\mathcal{C}^*} \;=\; \big([T]^{\mathcal{C}}_{\mathcal{B}}\big)^t, \tag{4.37}$$

where the ^{t} denotes the matrix transpose. The map $T \mapsto T^*$ gives a vector space isomorphism $\mathcal{L}(V,W) \cong \mathcal{L}(W^*,V^*)$. Let Y be another finite-dimensional F-vector space, and let $S \in \mathcal{L}(W,Y)$. Note that

$$(ST)^* \;=\; T^*S^* \quad \text{in} \quad \mathcal{L}(Y^*,V^*). \tag{4.38}$$

Thus, taking $W = V$, the map $\mathcal{L}(V) \to \mathcal{L}(V^*)$ given by $T \mapsto T^*$ is a *ring anti-homomorphism* since it is an additive group homomorphism but reverses the order of multiplication.

4.16. Let U be a subspace of the finite-dimensional F-vector space V. We have the canonical projection map $\pi \in \mathcal{L}(V, V/U)$ given by $v \mapsto v + U$, with dual map $\pi^* \in \mathcal{L}((V/U)^*, V^*)$. Also, there is the inclusion map $\iota \in \mathcal{L}(U,V)$ given by $u \mapsto u$, with dual map $\iota^* \in \mathcal{L}(V^*,U^*)$.

(i) Prove that π^* is injective and ι^* is surjective.

(ii) Let

$$U^\perp \;=\; \{\, y \in V^* \mid y(u) = 0, \text{ for all } u \in U\}. \tag{4.39}$$

Prove that

$$U^\perp \;=\; \mathit{im}(\pi^*) \;=\; \mathit{ker}(\iota^*).$$

Hence, from part (i),

$$U^\perp \cong (V/U)^* \quad \text{and} \quad V^*\big/U^\perp \cong U^*,$$

and from (4.35) and (4.8)

$$\mathit{dim}(U^\perp) \;=\; \mathit{dim}(V) - \mathit{dim}(U). \tag{4.40}$$

(iii) For any subspaces U and W of V, prove that

$$(U+W)^\perp \;=\; U^\perp \cap W^\perp \;\text{ and }\; (U \cap W)^\perp \;=\; U^\perp + W^\perp.$$

(iv) Let $y_1, \ldots, y_k \in V^*$ and let $Y = \mathit{span}\{y_1, \ldots, y_k\}$, a subspace of V^*. Let

$$W \;=\; \textstyle\bigcap_{i=1}^k \mathit{ker}(y_i) \;=\; \bigcap_{y \in Y} \mathit{ker}(y) \;\subseteq\; V.$$

Prove that $W^\perp = Y$.

(v) Prove that the map $U \mapsto U^\perp$ gives an inclusion-reversing one-to-one correspondence between the subspaces of V and the subspaces of V^*.

4.17. The *double dual* of the finite-dimensional vector space V is

$$V^{**} = (V^*)^* = \mathcal{L}(V^*, F). \tag{4.41}$$

(i) For every $v \in V$ there is a corresponding map $v^{**} \in V^{**}$ given by

$$v^{**}(y) = y(v), \quad \text{for all } y \in V^*.$$

Prove that the map $\zeta\colon V \to V^{**}$ given by $v \mapsto v^{**}$ is a vector space isomorphism. Thus, ζ gives a canonical isomorphism between V and V^{**}. By contrast, even though $V \cong V^*$ (since $dim(V^*) = dim(V)$) there is no natural choice of isomorphism between V and V^*.

(ii) Let U be any subspace of V. Then $U^\perp$ is a subspace of V^*, and there is a corresponding subspace $(U^\perp)^\perp \subseteq V^{**}$. Prove that $(U^\perp)^\perp = \zeta(U)$.

4.18. Let W be an infinite-dimensional F-vector space, and let $W^* = \mathcal{L}(W, F)$ be its dual space. Prove that $dim(W^*) > dim(W)$ as infinite cardinal numbers. It follows that the canonical injective map $W \to W^{**}$ is not surjective.

4.19. Let V and W be finite-dimensional F-vector spaces, and let $T \in \mathcal{L}(V, W)$.

(i) Prove that $im(T^*) = (ker(T))^\perp$ in V^*.

(ii) Deduce that

$$rk(T^*) = rk(T). \tag{4.42}$$

4.20. Let $A \in F^{m\times n}$. Prove that $row\text{-}rk(A) = col\text{-}rk(A)$ using the preceding problem.

4.21. Take any $T \in \mathcal{L}(V)$, and let W be a subspace of V.

(i) Prove that W is T-invariant iff $W^\perp$ is T^*-invariant.

(ii) Suppose that W is T-invariant, and let $\overline{T} \in \mathcal{L}(V/W)$ be the linear transformation induced by T. For $v \in V$, let $\overline{v} = v + W \in V/W$. Let $\mathcal{C} = \{w_1, \ldots, w_k\}$ be a base of W,

and choose $v_1, \dots, v_{n-k} \in V$, so that $\overline{\mathcal{D}} = \{\overline{v_1}, \dots, \overline{v_{n-k}}\}$ is a base of V/W. Thus, $\mathcal{B} = \{w_1, \dots, w_k, v_1, \dots, v_{n-k}\}$ is a base of V (see problem 4.7). Let $\mathcal{B}^* = \{w_1^* \dots, w_k^*, v_1^*, \dots, v_{n-k}^*\}$ be the corresponding dual base of V^*. Let $\mathcal{D}^* = \{v_1^*, \dots, v_{n-k}^*\}$. Prove that $\mathcal{D}^*$ is a base of $W^\perp$ and that

$$[T^*|_{W^\perp}]_{\mathcal{D}^*} \;=\; \big([\overline{T}]_{\overline{\mathcal{D}}}\big)^t.$$

(See (4.32) and (4.37).)

(iii) In the setting of part (ii), let $\overline{T^*} \in \mathcal{L}(V^*/W^\perp)$ be the linear transformation induced by $T^* \in \mathcal{L}(V^*)$. Let

$$\overline{w_i^*} \;=\; w_i^* + W^\perp \;\in\; V^*/W^\perp$$

for $i = 1, 2, \dots, k$, and let $\overline{\mathcal{C}^*} = \big\{\overline{w_1^*}, \dots, \overline{w_k^*}\big\}$. Prove that $\overline{\mathcal{C}^*}$ is a base of $V^*/W^\perp$ and that

$$[\overline{T^*}]_{\overline{\mathcal{C}^*}} \;=\; \big([T|_W]_{\mathcal{C}}\big)^t.$$

4.4. Determinants

Throughout this section, let R be a commutative ring.

Recall from (2.30) that for the symmetric group S_n, $n \geq 2$, we have the sign function $sgn\colon S_n \to \{1, -1\}$, which is a group homomorphism with kernel the alternating group A_n. (For $n = 1$, let sgn be the trivial homomorphism $S_1 \to \{1\}$.)

For any $n \in \mathbb{N}$, let $A = (a_{ij}) \in M_n(R)$. The *determinant* of A is defined to be

$$det(A) \;=\; \sum_{\sigma \in S_n} sgn(\sigma) a_{1\sigma(1)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)} \;\in\; R. \tag{4.43}$$

Properties of the determinant 4.22. We recall some of the basic properties of the determinant. Proofs of these formulas can be found in many linear algebra or abstract algebra texts, including Hoffman & Kunze [**8**]. Most of the properties follow readily from the definition of the determinant by direct calculation and facts about permutations; we give a proof of the crucial product formula in part (ix) below.

(i) *Row linearity.* Let $A = (a_{ij}) \in M_n(R)$. Let $\rho_1, \rho_2, \dots, \rho_n$ be the rows of A. That is, $\rho_i = (a_{i1}, a_{i2}, \dots, a_{in}) \in R^{1\times n}$ for $i = 1, 2, \dots, n$. For $r \in R$, $r\rho_i$ means $(ra_{i1}, ra_{i2}, \dots, ra_{in})$.

Take any $\rho' \in F^{1\times n}$ and any $r, s \in R$ and any $i \in \{1, 2, \ldots, n\}$. Then,

$$det\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_{i-1} \\ r\rho_i + s\rho' \\ \rho_{i+1} \\ \vdots \\ \rho_n \end{pmatrix} = r\, det\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_{i-1} \\ \rho_i \\ \rho_{i+1} \\ \vdots \\ \rho_n \end{pmatrix} + s\, det\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_{i-1} \\ \rho' \\ \rho_{i+1} \\ \vdots \\ \rho_n \end{pmatrix}.$$

Each of the matrices in the formula has same j-th row ρ_j for $j \neq i$. This is the row linearity of the determinant in the i-th row. The row linearity holds for each i.

(ii) For any $A \in M_n(R)$ and $r \in R$, $det(rA) = r^n\, det(A)$.

(iii) *Row rearrangements.* Let $A = (a_{ij}) \in M_n(R)$, and take any $\tau \in S_n$. Let $B = (b_{ij})$ where $b_{ij} = a_{\tau(i)j}$ for all i, j. That is, B is obtained from A by rearranging the order of the rows: the i-th row of B is the $\tau(i)$-th row of A. Then, $det(B) = sgn(\tau)\, det(A)$. In particular, if B is obtained from A by interchanging two rows, then $det(B) = -\, det(A)$, since $sgn(\tau) = -1$ for any transposition τ.

(iv) *Alternating property.* If any two rows of a matrix $A \in M_n(R)$ are the same, then $det(A) = 0$. (For, if the i-th and k-th rows of A are the same for $i \neq k$, take the transposition $\tau = (i\ k)$. Then, $S_n = A_n \cup A_n\tau$, a disjoint union, and for $\sigma \in A_n$, the σ and $\sigma\tau$ summands in (4.43) sum to 0.)

(v) *Transpose.* For any $A \in M_n(R)$,

$$det(A^t) = det(A),$$

where A^t is the transpose of A. It follows that the column versions of properties (i), (iii), and (iv) also hold.

(vi) *Expansion by minors.* For $A = (a_{ij}) \in M_n(R)$ with $n > 1$ and any $i, j \in \{1, 2, \ldots, n\}$, let $A(i|j) \in M_{n-1}(R)$ be the matrix obtained by deleting the i-th row and the j-th column from A. Then, for any fixed i,

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}\, det(A(i|j)). \tag{4.44}$$

This is called the formula for $det(A)$ by *expansion by minors* along the i-th row of A. There is an analogous formula

for $det(A)$ by expansion by minors down the j-column of A, for any j:

$$det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}\, det(A(i|j)). \tag{4.45}$$

(vii) *Triangular matrices.* If $A = (a_{ij}) \in M_n(R)$ is upper triangular, i.e., $a_{ij} = 0$ whenever $i > j$, then

$$det(A) = a_{11}a_{22}\dots a_{nn}.$$

This is also true when A is lower triangular, i.e., $a_{ij} = 0$ whenever $i < j$.

(viii) *Block triangular matrices.* Let $A \in M_n(R)$ have upper triangular block form

$$A = \begin{pmatrix} B & C \\ \mathbf{0} & D \end{pmatrix},$$

where for some $r \in \{1, 2, \dots, n-1\}$ we have $B \in M_r(R)$, $C \in R^{r\times(n-r)}$, $\mathbf{0}$ denotes the 0-matrix in $R^{(n-r)\times r}$, and $D \in M_{n-r}(R)$. Then,

$$det(A) = det(B) \cdot det(D).$$

The analogous result holds for matrices in lower triangular block form. By induction, the analogous result holds for block triangular matrices with more that two rows and columns of blocks.

(ix) *Product formula.* Take any A, B in $M_n(R)$. Then,

$$det(AB) = det(A) \cdot det(B). \tag{4.46}$$

Proof: Let $A = (a_{ij})$ and let $B = (b_{ij}) = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_n \end{pmatrix}$, where $\rho_i = (b_{i1}, \dots, b_{in})$ is the i-th row of B. Then,

$$\begin{aligned} &det(AB) \\ &= \sum_{\sigma\in S_n} sgn(\sigma) \big(\sum_{k_1=1}^{n} a_{1k_1} b_{k_1\sigma(1)} \big) \dots \big(\sum_{k_n=1}^{n} a_{nk_n} b_{k_n\sigma(n)} \big) \\ &= \sum_{k_1=1}^{n} \dots \sum_{k_n=1}^{n} a_{1k_1} \dots a_{nk_n}\, d_{k_1,k_2,\dots,k_n}, \end{aligned}$$

where

$$d_{k_1,k_2,\ldots,k_n} = \sum_{\sigma\in S_n} sgn(\sigma) b_{k_1\sigma(1)}\ldots b_{k_n\sigma(n)} = det\begin{pmatrix}\rho_{k_1}\\ \vdots \\ \rho_{k_n}\end{pmatrix}.$$

If $k_j = k_\ell$ with $j \neq \ell$, then $d_{k_1,k_2,\ldots,k_n} = 0$, by the alternating property (iv). On the other hand, $k_1,\ldots,k_n$ are all distinct iff there is $\tau \in S_n$ with $k_i = \tau(i)$ for all i; then,

$$d_{k_1,k_2,\ldots,k_n} = sgn(\tau)\, det(B)$$

by property (iii). Thus,

$$\begin{aligned} det(AB) &= \sum_{\tau\in S_n} a_{1\tau(1)}\ldots a_{n\tau(n)}\, sgn(\tau)\, det(B) \\ &= det(A)\cdot det(B). \end{aligned}$$

(x) If $P \in M_n(R)$ is invertible, then for any $A \in M_n(R)$,

$$det(P^{-1}AP) = det(A).$$

4.23. *Adjoint matrix.* Let $A = (a_{ij}) \in M_n(R)$. For $i, j \in \{1, 2, \ldots, n\}$, the ij*-cofactor* of A is

$$(-1)^{i+j}\, det(A(i|j))$$

(with notation as in Property 4.22(vi) above). The *classical adjoint* of A, denoted $adj(A)$ is the transpose of the matrix of cofactors of A. That is,

$$\begin{aligned} adj(A) &= (adj(A)_{ij}) \in M_n(R), \\ &\text{where} \quad adj(A)_{ij} = (-1)^{i+j}\, det(A(j|i). \end{aligned} \tag{4.47}$$

(i) Prove that

$$A\cdot adj(A) = det(A)I_n = adj(A)\cdot A. \tag{4.48}$$

($det(A)I_n$ is the diagonal matrix with every ii-entry $det(A)$ and all other entries 0.) (Hint: Relate the entries of $A\cdot adj(A)$ to the expansion by minors formulas in Property 4.22(vi) above for A and also for A with one row replaced by another row.)

(ii) Deduce that A has an inverse in $M_n(R)$ iff $det(A) \in R^*$ (the group of units of R). Indeed, when $det(A) \in R^*$,

$$A^{-1} = det(A)^{-1}\, adj(A). \tag{4.49}$$

4.24. Let S and R be commutative rings. A ring homomorphism $f\colon S \to R$ induces a ring homomorphism $\widetilde{f}\colon M_n(S) \to M_n(R)$ given by

$$\widetilde{f}\big((b_{ij})\big) \;=\; (f(b)_{ij}).$$

Clearly,

$$det(\widetilde{f}(B)) \;=\; f(det(B)) \quad \text{and} \quad adj(\widetilde{f}(B)) \;=\; \widetilde{f}(adj(B))$$

for any $B \in M_n(S)$. Now prove that for any $A \in M_n(R)$ there is an integral domain S, a homomorphism $f\colon S \to R$, and a matrix $C \in M_n(S)$ such that $det(C) \neq 0$ and $\widetilde{f}(C) = A$. (Hint: Consider iterated polynomial rings.)

4.25. Let $A, B \in M_n(R)$ for any $n \geq 2$.

(i) Prove that

$$adj(AB) \;=\; adj(B) \cdot adj(A).$$

(ii) Prove that

$$adj(adj(A)) \;=\; det(A)^{n-2} A.$$

(iii) Let P be any invertible matrix in $M_n(R)$. Prove that

$$adj(P^{-1}AP) \;=\; P^{-1}\, adj(A)\, P.$$

(Hint: For each of (i)–(iii), when A and B are invertible, use formula (4.48). When they are not invertible, use problem 4.24. Alternatively, you can deduce the noninvertible case from the invertible case using problem 3.51 and Note 3.52, since the entries of the matrices in the formulas are polynomial functions of the entries of A and B.)

4.26. *Vandermonde matrix.* Take any $c_1, c_2, \ldots, c_n \in R$. Prove that

$$det \begin{pmatrix} 1 & c_1 & c_1^2 & \ldots & c_1^j & \ldots & c_1^{n-1} \\ 1 & c_2 & c_2^2 & \ldots & c_2^j & \ldots & c_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & c_i & c_i^2 & \ldots & c_i^j & \ldots & c_i^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & c_n & c_n^2 & \ldots & c_n^j & \ldots & c_n^{n-1} \end{pmatrix} \;=\; \prod_{1 \leq i < j \leq n} (c_j - c_i). \tag{4.50}$$

The matrix here is called a *Vandermonde matrix*. Its ij-entry is c_i^{j-1}.

4.27. *Cramer's Rule.* Let A be an invertible matrix in $M_n(R)$, and let $b \in R^n$. Write $A = [\gamma_1, \gamma_2, \ldots, \gamma_n]$, where γ_j is the j-th column of A. Prove *Cramer's Rule*, which says that the unique solution $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ to the system of linear equations $Ax = b$ is given by

$$x_j \;=\; det(A)^{-1}\, det\big([\gamma_1, \ldots, \gamma_{j-1}, b, \gamma_{j+1}, \ldots, \gamma_n]\big).$$

4.28. *Determinantal rank.* Let $A \in F^{m\times n}$, where F is a field. For $r \in \mathbb{N}$ with $r \le min(m, n)$, an $r \times r$ submatrix of A is what is left after deleting $m - r$ rows and $n - r$ columns from A. The *determinantal rank* of A, denoted $det\text{-}rk(A)$, is defined to be the largest integer r such that A has an $r \times r$ submatrix with nonzero determinant. Prove that

$$det\text{-}rk(A) = col\text{-}rk(A) = row\text{-}rk(A).$$

(See problem 4.14 above for the second equality.)

4.29. For any $m, n \in \mathbb{N}$, determine $\big|\, GL_n(\mathbb{Z}_m)\big|$. (Hint: First do this for m is a prime power, and recall that the m prime case was given in (2.61).)

Trace. Let $A = (a_{ij}) \in M_n(R)$. The *trace* of A is the sum of its diagonal elements:

$$tr(A) \;=\; a_{11} + a_{22} + \ldots + a_{nn} \in R. \tag{4.51}$$

Note that for $A, B \in M_n(R)$ and $c \in R$,

$$tr(A+B) \;=\; tr(A)+tr(B), \;\; tr(cA) = c\,tr(A), \;\text{ and }\; tr(AB) = tr(BA). \tag{4.52}$$

Characteristic polynomial. Let $A = (a_{ij}) \in M_n(R)$. In $M_n(R[X])$, we have the matrix

$$X\cdot I_n - A \;=\; \begin{pmatrix} X - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & X - a_{22} & \ldots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \ldots & X - a_{nn} \end{pmatrix}.$$

The *characteristic polynomial* of A is defined to be

$$\chi_A \;=\; det(X\cdot I_n - A) \;\in R[X]. \tag{4.53}$$

Note that

$$\chi_A = \prod_{i=1}^{n}(X - a_{ii}) + \text{(summands of degree at most } n-2);$$

hence,

$$\chi_A = X^n + (-tr(A))X^{n-1} + \ldots + (-1)^n\, det(A).$$

Example 4.30.

(i) If $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(R)$, then

$$\chi_A = X^2 - (a+d)X + (ad - bc).$$

(ii) If $A = (a_{ij}) \in M_n(R)$ is upper or lower triangular, then

$$\chi_A = (X - a_{11})(X - a_{22}) \ldots (X - a_{nn}).$$

(iii) For any $A \in M_n(R)$, $\chi_{A^t} = \chi_A$.

(iv) For any $A, P \in M_n(R)$ with P invertible, $\chi_{P^{-1}AP} = \chi_A$.

Let V be a finite-dimensional F-vector space, and let $T \in \mathcal{L}(V)$. Take any base $\mathcal{B}$ of V, and let $A = [T]_{\mathcal{B}}$. Define the determinant, trace, and characteristic polynomial of T by

$$det(T) = det(A), \quad tr(T) = tr(A), \quad \text{and} \quad \chi_T = \chi_A. \tag{4.54}$$

These are all well-defined independent of the choice of $\mathcal{B}$ because of the invariance under similarity of the determinant (see Property 4.22(x)), the trace (see (4.52)) and the characteristic polynomial (see Example 4.30(iv)).

4.31. Let $T \in \mathcal{L}(V)$ for a finite-dimensional F-vector space V, and let W be a T-invariant subspace of V. Recall the linear transformations $T|_W \in \mathcal{L}(W)$ and $\overline{T} \in \mathcal{L}(V/W)$ induced by T, as in (4.32). Prove that

$$\begin{gathered} det(T) = det(T|_W) \cdot det(\overline{T}), \quad tr(T) = tr(T|_W) + tr(\overline{T}), \\ \text{and} \quad \chi_T = \chi_{T|_W} \cdot \chi_{\overline{T}}. \end{gathered} \tag{4.55}$$

4.32. Let $A, B \in M_n(R)$ prove that $\chi_{AB} = \chi_{BA}$. (Hint: If A or B is invertible, this follows from Example 4.30(iv). If not, use problem 4.24.)

4.33. Let $A \in R^{m\times n}$ and $B \in R^{n\times m}$ with $m < n$. Prove that

$$\chi_{BA} = X^{n-m}\chi_{AB}.$$

4.34. For any field F and any $n \geq 2$, let A be the tridiagonal matrix in $M_n(F)$

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 1 & \ddots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots & 1 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix},$$

with all 0's on the main diagonal, all 1's on the first superdiagonal and subdiagonal, and 0's elsewhere. Determine χ_A.

4.35. Let

$$A = \begin{pmatrix} 1 & 1/2 & 1/3 & \dots & 1/n \\ 1/2 & 1/3 & 1/4 & \dots & 1/(n+1) \\ 1/3 & 1/4 & 1/5 & \dots & 1/(n+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/n & 1/(n+1) & 1/(n+2) & \dots & 1/(2n-1) \end{pmatrix} \in M_n(\mathbb{Q}).$$

The ij-entry of A is $1/(i+j-1)$. Determine $det(A)$, which is nonzero, and show that $A^{-1} \in M_n(\mathbb{Z})$. (Hint: For $det(A)$, more generally compute $det\big((c_{ij})\big)$, where $c_{ij} = 1/(a_i + b_j)$ for some $a_1, \dots, a_n, b_1, \dots, b_n$ in a field F, where $a_i + b_j \neq 0$ for all i, j.)

4.36. Let $A, B \in M_n(\mathbb{R})$. Suppose that A and B are similar in $M_n(\mathbb{C})$. Prove that they are already similar in $M_n(\mathbb{R})$. (This is an easy consequence of the uniqueness of the rational canonical form for A, see §4.9 below. But this problem asks you to give a proof without using canonical forms.)

4.5. Eigenvalues and eigenvectors, triangulation and diagonalization

Throughout this section, let V be a finite-dimensional F-vector space.

Let $T \in \mathcal{L}(V)$. An element $\lambda \in F$ is an *eigenvalue* of T if there is a nonzero $v \in V$ such that

$$T(v) = \lambda v.$$

Such a vector v is called an *eigenvector* of T for the eigenvalue λ. Analogously, for $B \in M_n(F)$, λ is an eigenvalue of B if $Bu = \lambda u$ for some $u \in F^n \setminus \{0\}$, and any such u is an eigenvector of B for λ. Back to T, let $A = [T]_{\mathcal{B}} \in M_n(F)$ for some base $\mathcal{B}$ of F. Since $[\lambda\, id_V - T]_{\mathcal{B}} = \lambda I_n - A$, we have λ is an eigenvalue of T iff $ker(\lambda\, id_V - T)$ is nontrivial, iff $\lambda I_n - A$ is not invertible, iff

$$0 = det(\lambda\, id_V - T) = det(\lambda I_n - A) = \chi_A(\lambda) = \chi_T(\lambda).$$

Thus,

$$\lambda \text{ is an eigenvalue of } T \quad \text{iff} \quad \lambda \text{ is a root of } \chi_T. \tag{4.56}$$

Hence, T has at most $n = deg(\chi_T) = dim(V)$ different eigenvalues (but possibly none at all if χ_T has no factors of degree 1 in $F[X]$).

Example 4.37. Let $T \in \mathcal{L}(V)$.

(i) If $[T]_{\mathcal{B}}$ is a triangular matrix, the diagonal entries of $[T]_{\mathcal{B}}$ are the eigenvalues of T (see Example 4.30(ii)).

(ii) If W is a T-invariant subspace of V and $T|_W \in \mathcal{L}(W)$ and $\overline{T} \in \mathcal{L}(V/W)$ are the associated linear transformations, then λ is an eigenvalue of T iff λ is an eigenvalue of $T|_W$ or of $\overline{T}$ (see (4.55)).

If λ is an eigenvalue of $T \in \mathcal{L}(V)$ the associated *λ-eigenspace of T* is

$$V_\lambda = \{v \in V \mid T(v) = \lambda v\} = ker(T - \lambda\, id_V), \tag{4.57}$$

which consists of 0_V together with all the λ-eigenvectors of T. Note that V_λ and each of its subspaces is T-invariant. A short calculation shows that a family of eigenvectors of T all having different eigenvalues is linearly independent. Hence, if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct

eigenvalues of T, then

$$V_{\lambda_1} + V_{\lambda_2} + \ldots + V_{\lambda_k} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \ldots \oplus V_{\lambda_k} \quad \text{in } V. \tag{4.58}$$

4.38. Let $A \in M_n(\mathbb{R})$. Then, A may have no eigenvalues in $\mathbb{R}$, since χ_A may not have any roots in $\mathbb{R}$. However, if we view A in $M_n(\mathbb{C})$, then χ_A is unchanged, and it has a root in $\mathbb{C}$ as $\mathbb{C}$ is algebraically closed (see problem 5.101 below). Therefore, A has an eigenvector $v \in \mathbb{C}^n$, and v can be written uniquely as $v = v_1 + iv_2$, with $v_1, v_2 \in \mathbb{R}^n$. Prove that the $\mathbb{R}$-subspace $span\{v_1, v_2\} \subseteq \mathbb{R}^n$ is A-invariant (i.e., L_A-invariant for the multiplication-by-A linear transformation $L_A \in \mathcal{L}(\mathbb{R}^n)$). Thus, any $A \in M_n(\mathbb{R})$ has an A-invariant subspace of $\mathbb{R}^n$ of dimension 1 or 2.

Split polynomials. A polynomial $f \in F[X]$ with $deg(f) \geq 1$ is said to *split* over F if f is a product of polynomials of degree 1 in $F[X]$. It is a useful consequence of Kronecker's theorem on roots of polynomials (see Note 5.2 below) that

$$\begin{array}{l}\text{even if } f \text{ does not split over } F, \text{ there is} \\ \text{a field } K \supseteq F \text{ such that } f \text{ splits over } K.\end{array} \tag{4.59}$$

Triangulable linear transformations and matrices. A linear transformation $T \in \mathcal{L}(V)$ is said to be *triangulable* if there is a base $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is an upper triangular matrix. A matrix $A \in M_n(F)$ is said to be triangulable if A is similar to an upper triangular matrix in $M_n(F)$. Thus, T is triangulable iff $[T]_{\mathcal{C}}$ triangulable for any base $\mathcal{C}$ of V. We now show that

$$T \text{ is triangulable} \quad \text{iff} \quad \chi_T \text{ is split over } F. \tag{4.60}$$

Proof: If T is triangulable, then χ_T is split, by Example 4.30(ii). For the converse, suppose that χ_T is split over F, and argue by induction on $n = dim(V)$. The case $n = 1$ is clear; assume that $n > 1$. Since χ_T is split it has a root $\lambda \in F$, which is an eigenvalue of T. Let v be a λ-eigenvector of T, and let $W = span\{v\}$, a 1-dimensional T-invariant subspace of V. For the induced linear transformation $\overline{T} \in \mathcal{L}(V/W)$, we have $\chi_{\overline{T}}$ is split, since it divides χ_T (see (4.55)). Hence, by induction as $dim(V/W) < n$, there is a base $\overline{\mathcal{B}} = \{\overline{v_2}, \ldots \overline{v_n}\}$ of V/W with $[\overline{T}]_{\overline{\mathcal{B}}}$ upper triangular. Then for the base $\mathcal{B} = \{v, v_2, \ldots, v_n\}$ of V, where $v_i + W = \overline{v_i}$, the matrix $[T]_{\mathcal{B}}$ is upper triangular by (4.32).

It follows immediately from (4.60) and (4.55) that if W is a T-invariant subspace of $T \in \mathcal{L}(V)$, with induced linear transformations $T|_W \in \mathcal{L}(W)$ and $\overline{T} \in \mathcal{L}(V/W)$, then T is triangulable iff $T|_W$ and $\overline{T}$ are each triangulable.

4.39. Let $T \in \mathcal{L}(V)$ Prove that if $\chi_T = (X - \lambda_1) \dots (X - \lambda_n)$, then for any $f \in F[X]$,

$$\chi_{f(T)} = (X - f(\lambda_1)) \dots (X - f(\lambda_n)).$$

4.40. *Simultaneous triangulability.* Let, $T, S \in \mathcal{L}(V)$, and suppose that $ST = TS$.

(i) Let λ be an eigenvalue of T, with corresponding eigenspace V_λ. Prove that V_λ is S-invariant.

(ii) Suppose that T and S are each triangulable. Then T has an eigenvalue λ. Since the eigenspace V_λ of T for λ is S-invariant and $S|_{V_\lambda}$ is triangulable by (4.60) as $\chi_{S|_{V_\lambda}} \big| \chi_S$, there is an eigenvector v for S lying in V_λ. Use this to prove that there is a base $\mathcal{B}$ of V such that $[T]_\mathcal{B}$ and $[S]_\mathcal{B}$ are each upper triangular. We say that T and S are *simultaneously triangulable.*

(iii) Now generalize the result of part (ii): Take $T_1, T_2, \dots, T_k$ in $\mathcal{L}(V)$, and suppose that $T_i T_j = T_j T_i$ for all i, j. Prove that there is a base $\mathcal{B}$ of V such that each of $[T_1]_\mathcal{B}$, $[T_2]_\mathcal{B}$, $\dots$, $[T_k]_\mathcal{B}$ is upper triangular.

4.41. Let A be an invertible matrix in $M_n(F)$, and let

$$\chi_A = X^n + c_{n-1}X^{n-1} + \dots + c_1 X + c_0.$$

Then, $c_0 = (-1)^n \, det(A) \neq 0$. Prove that

$$\chi_{A^{-1}} = c_0^{-1}\big[c_0 X^n + c_1 X^{n-1} + \dots + c_{n-i}X^i + \dots + c_{n-1}X + 1\big].$$

(Hint: Prove this first for A triangular.)

Diagonal matrices. A matrix $A = (a_{ij}) \in M_n(F)$ is said to be *diagonal* if $a_{ij} = 0$ whenever $i \neq j$. That is, the only nonzero entries of A occur on its main diagonal. When A is diagonal, we write

$$A = diag(a_{11}, a_{22}, \dots, a_{nn}). \tag{4.61}$$

Note that the set of all diagonal matrices in $M_n(F)$ is an F-vector subspace and a subring of $M_n(F)$ that is ring isomorphic to $\prod_{i=1}^{n} F$.

Diagonalizable linear transformations. A linear transformation $T \in \mathcal{L}(V)$ is said to be *diagonalizable* if there is a base $\mathcal{B}$ of F such that $[T]_{\mathcal{B}}$ is a diagonal matrix. Note that when $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $[T]_{\mathcal{B}} = diag(\lambda_1, \ldots, \lambda_n)$, then each v_i is an eigenvalue of T for the eigenvector λ_i. Thus, T is diagonalizable iff V has a base consisting of eigenvectors of T. Hence, if $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T with corresponding eigenspaces $V_{\lambda_1}, \ldots, V_{\lambda_k}$, then

$$T \text{ is diagonalizable iff } V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}. \tag{4.62}$$

In particular, if T has $n = dim(V)$ distinct eigenvalues, then T is diagonalizable. A matrix $B \in M_n(F)$ is said to be diagonalizable if B is similar to a diagonal matrix. Thus, if $\mathcal{B}$ is any base of V, then T in $\mathcal{L}(V)$ is diagonalizable iff $[T]_{\mathcal{B}}$ in $M_n(F)$ is diagonalizable. Now take an invertible $P \in M_n(F)$. Note that $P^{-1}BP = diag(\lambda_1, \ldots, \lambda_n)$ iff the j-th column of P is a λ_j-eigenvector for B.

4.42. *Nilpotent linear transformations.* A linear transformation N in $\mathcal{L}(V)$ is said to be *nilpotent* if $N^k = 0$ for some $k \in \mathbb{N}$.

(i) Suppose $N \in \mathcal{L}(V)$. Prove that the following conditions are equivalent:
  (a) N is nilpotent.
  (b) N is triangulable and 0 is its only eigenvalue.
  (c) $\chi_N = X^n \in F[X]$, where $n = dim(V)$.
  (d) There is a base $\mathcal{B}$ of V such that $[N]_{\mathcal{B}}$ is *strictly upper triangular*, i.e., upper triangular with all 0's on the main diagonal.
  (e) There are subspaces $W_0, W_1, \ldots, W_n$ of V such that $W_0 \subseteq W_1 \subseteq \ldots \subseteq W_n$ with $dim(W_i) = i$ for all i and $N(W_i) \subseteq W_{i-1}$ for $i \geq 1$.

(ii) If N is nilpotent, prove that N is diagonalizable iff $N = 0$.

(iii) If N is nilpotent, prove that $N^n = 0$, where $n = dim(V)$.

4.43. Let $A \in M_n(F)$. We say that A is *nilpotent* if $A^k = 0$ for some $k \in \mathbb{N}$.

(i) Prove that if A is nilpotent, then $tr(A^j) = 0$ for every $j \in \mathbb{N}$.

(ii) Conversely, prove that if $char(F) = 0$ or $char(F) > n$ and $tr(A) = tr(A^2) = ... = tr(A^n) = 0$, prove that A is nilpotent. (The assumption on $char(F)$ is needed here: For any prime number p and field F with $char(F) = p$, consider the identity matrix in $M_p(F)$.)

4.44. *Fibonacci sequence.* Let $f_1, f_2, \ldots$ be the Fibonacci sequence, defined recursively by $f_1 = 1$, $f_2 = 2$, and $f_{i+1} = f_i + f_{i-1}$ for $i \geq 2$, as in problem 1.1. For convenience, set $f_0 = 0$.

(i) Find a matrix $A \in M_2(\mathbb{R})$ such that $\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = A \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix}$ for every $n \in \mathbb{N}$. It then follows by induction that

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = A^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \quad \text{for all} \quad n \in \mathbb{N}.$$

(ii) Determine the eigenvalues of A, and use this to find an invertible matrix $P \in M_2(\mathbb{R})$ such that $D = P^{-1}AP$ is a diagonal matrix.

(iii) Use the information from part (ii) to find explicit formulas for the entries of A^n, for each $n \in \mathbb{N}$. (Note that $A^n = PD^nP^{-1}$.) Use this to deduce the closed formula for f_n given in (1.2).

4.45. A matrix $A = (a_{ij})$ in $M_n(F)$ is *symmetric* if $A = A^t$, i.e., $a_{ij} = a_{ji}$ for all i, j.

(i) Suppose that $char(F) \neq 2$. Let $F^2 = \{c^2 \mid c \in F\}$. Prove that every symmetric matrix in $M_2(F)$ is diagonalizable iff every sum of two squares in F is already a square in F (i.e., F^2 is closed under addition) and $-1 \notin F^2$. (Note that since $char(F) \neq 2$ the quadratic formula holds for roots of polynomials of degree 2 in $F[X]$.) This holds in particular for $F = \mathbb{R}$; more generally in fact, every symmetric matrix in $M_n(\mathbb{R})$ is diagonalizable, for every $n \in \mathbb{N}$; see the comments preceding problem 4.105 below.

(ii) Prove that if $char(F) = 2$, then there is a symmetric matrix in $M_2(F)$ that is not diagonalizable.

4.46. While symmetric matrices over $\mathbb{R}$ are diagonalizable, symmetric matrices over $\mathbb{C}$ need not be diagonalizable. For diagonalizability, the

right generalization of real symmetric matrices is Hermitian matrices over $\mathbb{C}$. A matrix $A = (a_{ij}) \in M_n(\mathbb{C})$ is said to be *Hermitian* if $a_{ji} = \overline{a_{ij}}$ for all i, j, where the bar denotes complex conjugate. Prove that if $A \in M_2(\mathbb{C})$ is Hermitian, then A is diagonalizable. (It is in fact true that Hermitian matrices in $M_n(\mathbb{C})$ are diagonalizable for every $n \in \mathbb{N}$; see, e.g., Hoffman & Kunze, [**8**, p. 314]).

4.47. Let $A_1, A_2, A_3, A_4 \in M_n(F)$, and suppose that $A_i A_j = A_j A_i$ for all i, j. Prove that

$$det \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} = det(A_1 A_4 - A_3 A_2).$$

(Note that by passing from F to a larger field and using (4.59), one can assume that the A_i are each triangulable.)

4.6. Minimal polynomials of a linear transformation and primary decomposition

Throughout this section V is a finite-dimensional F-vector space with $dim(V) = n$.

Evaluation map and minimal polynomial. Let $T \in \mathcal{L}(V)$, and let $f = c_0 + c_1 X + \ldots + c_k X^k \in F[X]$ (with the $c_i \in F$). Then define

$$f(T) = c_0\, id_V + c_1 T + \ldots + c_k T^k \in \mathcal{L}(V). \tag{4.63}$$

There is an associated *evaluation at* T map ε_T: $F[X] \to \mathcal{L}(V)$ given by

$$\varepsilon_T(f) = f(T).$$

Note that ε_T is a ring and F-vector space homomorphism. Let

$$F[T] = im(\varepsilon_T) = \{f(T) \mid f \in F[X]\}. \tag{4.64}$$

Thus, $F[T]$ is a subring and F-subspace of $\mathcal{L}(V)$. By the FHT,

$$F[T] \cong F[X]/\ker(\varepsilon_T),$$

a ring and vector space isomorphism. Since $dim(F[X]) = \infty$ but $dim(F[T]) \leq dim(\mathcal{L}(V)) < \infty$, the ideal $ker(\varepsilon_T)$ of $F[X]$ is nontrivial. As $F[X]$ is a PID, $ker(\varepsilon_T)$ is a principal ideal; hence, $ker(\varepsilon_T)$ has a generator, which is unique up to multiplication by an element in $F[X]^* = F^*$. We choose the unique monic generator, and call it the

minimal polynomial of T, denoted m_T. Thus, m_T is characterized by each of the following conditions:

(a) m_T is monic in $F[X]$, and for $f \in F[X]$, we have $f(T) = 0$ iff $m_T | f$;

(b) m_T is the monic polynomial in $F[X]$ of least degree such that $m_T(T) = 0$.

Let $d = deg(m_T) \geq 1$, and let $\mathcal{B} = \{id_V, T, T^2, \ldots, T^i, \ldots, T^{d-1}\}$. Since the polynomials in $ker(\varepsilon_T)$ correspond to linear dependence relations among the powers of T, the elements of $\mathcal{B}$ are linearly independent. They also span $F[T]$, as one can see using the Division Algorithm in $F[X]$ (dividing m_T into a polynomial). Thus, $\mathcal{B}$ is a base of $F[T]$, and

$$dim(F[T]) \;=\; deg(m_T). \tag{4.65}$$

Note also that since $F[T] \cong F[X]/(m_T)$, the ring structure of the commutative ring $F[T]$ is determined by the prime factorization of m_T in $F[X]$.

If $B \in M_n(F)$, and $f = c_0 + c_1 X + \ldots + c_k X^k \in F[X]$, define

$$f(B) \;=\; c_0 I_n + c_1 X + \ldots + c_k B^k \;\in\; M_n(F). \tag{4.66}$$

Just as for linear transformations, there is an evaluation at B ring and vector space homomorphism $\varepsilon_B \colon F[X] \to M_n(F)$ given by

$$\varepsilon_B(f) \;=\; f(B),$$

whose image is denoted by $F[B]$, which is a commutative subring and vector subspace of $M_n(F)$. The minimal polynomial of B, denoted m_B, is the monic generator of the principal ideal $ker(\varepsilon_B)$ of $F[X]$. Then, $dim(F[B]) = deg(m_B)$. Note that if $T \in \mathcal{L}(V)$ and $A = [T]_{\mathcal{B}}$ for some base $\mathcal{B}$ of V, then $[f(T)]_{\mathcal{B}} = f(A)$ for every $f \in F[X]$. Hence, $m_T = m_A$ and the map $F[T] \to F[A]$ given by $f(T) \mapsto f(A)$ for $f \in F[X]$ is a well-defined ring and vector space isomorphism.

Example 4.48. Take any $T \in End(V)$, and let $T^* \in \mathcal{L}(V^*)$ be its dual linear transformation as in (4.36). For any $f \in F[X]$, we have $f(T^*) = f(T)^*$ since the map $\mathcal{L}(V) \to \mathcal{L}(V^*)$ given by $S \mapsto S^*$ is a linear transformation and ring anti-homomorphism (see (4.38)).

Hence, $m_{T^*} = m_T$. Also, $\chi_{T^*} = \chi_T$ by (4.37) and Property 4.22(v). Likewise, for $A \in M_n(F)$, $m_{A^t} = m_A$ and $\chi_{A^t} = \chi_A$.

4.49. Let $T \in \mathcal{L}(V)$, and suppose that m_T is an irreducible polynomial in $F[X]$.

(i) Prove that $F[T]$ is a field, and that V is an $F[T]$-vector space with its given addition and with scalar multiplication defined by $S \cdot v = S(v)$ for all $S \in F[T]$ and $v \in V$.

(ii) Prove that an F-subspace W of V is T-invariant iff W is an $F[T]$-subspace of V.

4.50. Let $T \in \mathcal{L}(V)$. Prove that $\lambda \in F$ is a root of m_T iff λ is an eigenvalue of T.

4.51. Take $T \in \mathcal{L}(V)$ and $g \in F[X]$. Prove that $g(T)$ is invertible in $\mathcal{L}(V)$ iff $gcd(g, m_T) \sim 1$ in $F[X]$.

4.52. Fix matrices $A, B \in M_n(F)$. Define $T \in \mathcal{L}(M_n(F))$ by

$$T(C) \;=\; AC - CB \quad \text{for all} \;\; C \in M_n(F).$$

Prove that T is invertible iff $gcd(m_A, m_B) \sim 1$. (Hint: If $AC = CB$, then $f(A)C = Cf(B)$ for every $f \in F[X]$.) (Note that when A and B are triangulable, the condition on the minimal polynomials is equivalent to: A and B have no common eigenvalue.)

4.53. Let $T \in \mathcal{L}(V)$, let W be a T-invariant subspace of V, and let $T|_W \in \mathcal{L}(W)$ and $\overline{T} \in \mathcal{L}(W/V)$ be the linear transformations induced by T. Note that W is $f(T)$-invariant for every $f \in F[X]$, and that $f(T|_W) = f(T)|_W$ and $f\big(\overline{T}\big) = \overline{f(T)}$. Moreover, there are well-defined ring and F-vector space homomorphisms $F[T] \to F[T|_W]$ (given by $f(T) \mapsto f(T|_W)$) and $F[T] \to F\big[\,\overline{T}\,\big]$ (given by $f(T) \mapsto f\big(\overline{T}\big)$).

(i) Prove that

$$m_{T|_W} \,|\, m_T \quad \text{and} \quad m_{\overline{T}} \,|\, m_T$$

in $F[X]$. Hence, $lcm(m_{T|_W}, m_{\overline{T}}) \,|\, m_T$.

(ii) For $g, h \in F[X]$, prove that if $h(\overline{T}) = 0$ in $\mathcal{L}(V/W)$ and $g(T|_W) = 0$ in $\mathcal{L}(W)$, then $(gh)(T) = 0$ in $\mathcal{L}(V)$.

(iii) Prove that $m_T \,|\, (m_{T|_W} \cdot m_{\overline{T}})$ in $F[X]$.

4.54. Let $T \in \mathcal{L}(V)$, and suppose that $V = W_1 \oplus W_2$, where W_1 and W_2 are each T-invariant. Prove that

$$m_T \;=\; lcm\left(m_{T|_{W_1}}, m_{T|_{W_2}}\right) \quad \text{and} \quad \chi_T \;=\; \chi_{T|_{W_1}} \cdot \chi_{T|_{W_2}}. \tag{4.67}$$

4.55. Let $T \in \mathcal{L}(V)$. Take any nonzero $g, h \in F[X]$, with $gcd(g,h) \sim 1$. By problem 3.56, there exist $k, \ell \in F[X]$ with $kg + \ell h = 1$. Suppose that $(gh)(T) = 0$ in $\mathcal{L}(V)$. In V, let $W_1 = ker(g(T))$ and $W_2 = ker(h(T))$.

(i) Prove that W_1 and W_2 are T-invariant, and that

$$V \;=\; W_1 \oplus W_2.$$

(ii) Prove that $W_1 = im(h(T))$ and $W_2 = im(g(T))$.

4.56. *Primary decomposition.* Let $T \in \mathcal{L}(V)$. The monic polynomial m_T has a unique factorization

$$m_T \;=\; q_1^{r_1} q_2^{r_2} \ldots q_k^{r_k} \;\text{ in }\; F[X],$$

where the q_i are distinct monic irreducible polynomials in $F[X]$ and the $r_i \in \mathbb{N}$. Let

$$U_i \;=\; ker\left(q_i^{r_i}(T)\right) \qquad \text{for } i = 1, 2, \ldots, k. \tag{4.68}$$

(i) Prove that each U_i is T-invariant and that

$$V \;=\; U_1 \oplus U_2 \oplus \ldots \oplus U_k. \tag{4.69}$$

(Argue by induction on k, using the preceding problem for the induction step.)

(ii) Prove that for each i,

$$m_{T|_{U_i}} \;=\; q_i^{r_i} \;\text{ and }\; U_i \;=\; im\left((q_1^{r_1} \ldots q_{i-1}^{r_{i-1}} q_{i+1}^{r_{i+1}} \ldots q_k^{r_k})(T)\right).$$

(iii) Prove that for each i,

$$U_i \;=\; \{v \in V \mid q_i^s(T)(v) = 0 \text{ for some } s \in \mathbb{N}\}. \tag{4.70}$$

The direct sum decomposition of V in (4.69) is called the *primary decomposition* of V relative to T, and the subspace U_i is called the *q_i-primary component* of V for T.

Example 4.57. Let $T \in \mathcal{L}(V)$, and let $\lambda_1, \ldots, \lambda_k$ be all the distinct eigenvalues of T.

(i) Suppose that T is diagonalizable with eigenspace decomposition $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$ as in (4.62). Since clearly

$$m_{T|_{V_{\lambda_i}}} = X - \lambda_i,$$

problem 4.54 (and induction) shows that

$$m_T = (X - \lambda_1)(X - \lambda_2) \ldots (X - \lambda_k).$$

Since the λ_i are distinct, the $q_i^{r_i}$ in the irreducible factorization of m_T in problem 4.56 are just the $X - \lambda_i$. Thus, the primary components U_i of V for T are the eigenspaces V_{λ_i}, and the primary decomposition of V for T coincides with the eigenspace decomposition noted in (4.62).

(ii) Conversely to part (i), suppose that m_T factors in $F[X]$ into a product of distinct monic polynomials of degree 1, i.e.,

$$m_T = (X - \mu_1)(X - \mu_2) \ldots (X - \mu_\ell),$$

with the μ_i distinct. Then, the $q_i^{r_i}$ in the irreducible factorization of m_T are the $X - \mu_i$ and the primary components U_i of V for T are just the T-eigenspaces V_{μ_i}. Thus, the primary decomposition of V is an eigenspace decomposition, so by (4.62), T is diagonalizable. Moreover, $\mu_1, \ldots, \mu_\ell$ are the eigenvalues of T, as noted already in problem 4.50.

(iii) It follows by parts (i) and (ii) that T is diagonalizable iff m_T is a product of distinct monic irreducibles of degree 1 in $F[X]$.

(iv) It follows from part (iii) and problem 4.53(i) that if T is diagonalizable and W is any T-invariant subspace of V, then the associated maps $T|_W \in \mathcal{L}(W)$ and $\overline{T} \in \mathcal{L}(V/W)$ are also diagonalizable.

Cayley–Hamilton Theorem. Let $T \in \mathcal{L}(V)$. The Cayley–Hamilton Theorem says that $m_T | \chi_T$ in $F[X]$ or, equivalently, $\chi_T(T) = 0$ in $\mathcal{L}(V)$. This was proved for T nilpotent in problem 4.42(i) and (iii). The next problem asks you to prove the theorem for T triangulable, and the general case is given in problem 4.68 below.

4.58. Let $T \in \mathcal{L}(V)$ with T is triangulable. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then,

$$\chi_T \;=\; (X-\lambda_1)^{s_1} \ldots (X-\lambda_k)^{s_k}, \tag{4.71}$$

for some $s_1, \ldots, s_k \in \mathbb{N}$.

(i) Let q be an irreducible monic factor of m_T in $F[X]$ and let U be the q-primary component of V for T. Recall from problem 4.56 that U is T-invariant and $m_{T|_U} = q^r$ for some $r \in \mathbb{N}$; also, $T|_U$ is triangulable by (4.60). Prove that $q = X - \mu$ for some eigenvalue μ of $T|_U$. Deduce that $q = X - \lambda_j$ for some j.

(ii) Problem 4.56 shows that

$$\begin{aligned} U \;&=\; ker((q^r)(T)) \;=\; ker\left((T-\lambda_j\, id_V)^r\right) \\ &= \{v \in V \mid (T-\lambda_j\, id_V)^m(v) = 0 \text{ for some } m \in \mathbb{N}\}, \end{aligned} \tag{4.72}$$

which is called the *generalized λ_j-eigenspace* of T. Note that as $T|_U$ is triangulable and λ_j is its only eigenvalue (since λ_j is the only root of $m_{T|_U}$),

$$\chi_{T|_U} \;=\; (X-\lambda_j)^t, \qquad \text{for some } t \in \mathbb{N}.$$

Let $N = T|_U - \lambda_j\, id_U \in \mathcal{L}(U)$. Prove that N is nilpotent, determine m_N and χ_N, and deduce that $r \leq t$. (Recall problem 4.42.)

(iii) It follows from part(i) that

$$m_T \;=\; (X-\lambda_1)^{r_1} \ldots (X-\lambda_k)^{r_k}$$

for some nonnegative integers $r_1, \ldots, r_k$. In fact, each $r_i \geq 1$, since λ_i is a root of χ_T, hence an eigenvalue of T, hence a root of m_T. Let U_i be the $(X-\lambda_i)$-primary component of V for T. By part (ii), U_i is the generalized λ_i-eigenspace of T and $\chi_{T_{U_i}} = (X-\lambda_i)^{t_i}$ for some $t_i \geq r_i$. Prove that each $t_i = s_i$ for the s_i of (4.71). (Recall (4.67).) It follows that $m_T | \chi_T$, proving the Cayley–Hamilton Theorem for triangulable T.

4.59. Let $T \in \mathcal{L}(V)$, and let $V = U_1 \oplus \ldots \oplus U_k$ be the primary decomposition of V for T, as in problem 4.56. Let W be any T-invariant subspace of V. Prove that

$$W = (U_1 \cap W) \oplus \ldots \oplus (U_k \cap W)$$

is the primary decomposition of W for $T|_W$ (after eliminating any trivial summands where $U_i \cap W = \{0\}$).

4.60. Let $T \in \mathcal{L}(V)$. Prove that

$$V = ker(T) \oplus im(T) \quad \text{iff} \quad X^2 \nmid m_T \text{ in } F[X].$$

4.61. *Simultaneous diagonalizability.* Let $T, S \in \mathcal{L}(V)$ with $ST = TS$.

(i) Prove that the primary components U_i of V for T are S-invariant.

(ii) Suppose that T and S are each diagonalizable. Prove that there is a base $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ and $[S]_{\mathcal{B}}$ are each diagonal matrices.

4.62. Let $T_1, \ldots, T_n \in \mathcal{L}(V)$ and let $T_1^*, \ldots, T_n^* \in \mathcal{L}(V^*)$ be their duals. Suppose that $T_iT_j = T_jT_i$ for all i, j.

(i) Supppose that λ_i is an eigenvalue for T_i, and that there is a *simultaneous λ_i-eigenvector* v for all the T_i, i.e., $v \in V \setminus \{0\}$ and $T_i(v) = \lambda_i v$ for each i. Prove that there is $y \in V^* \setminus \{0\}$ with $T_i^*(y) = \lambda_i y$ for each i.

(ii) Give an example to show that the result of part (i) is not true if we do not assume that $T_iT_j = T_jT_i$ for all i, j.

4.7. T-cyclic subspaces and T-annihilators

4.63. Let

$$f = X^n + c_{n-1}X^{n-1} + \ldots + c_1X + c_0$$

be any monic polynomial in $F[X]$ of positive degree. The *companion matrix* of f is

$$C_f = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & 0 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -c_{n-1} \end{pmatrix} \in M_n(F). \qquad (4.73)$$

All the entries in the first $n-1$ columns are 0's, except for all 1's on the first subdiagonal. (If $deg(f) = 1$, then $f = X + c_0$ and we set $C_f = (-c_0) \in M_1(F)$.) Prove that

$$\chi_{C_f} = f.$$

It follows from this or by direct calculation that

$$det(C_f) = (-1)^n c_0 \qquad \text{and} \qquad tr(C_f) = -c_{n-1}. \tag{4.74}$$

T-cyclic subspaces. Fix $T \in \mathcal{L}(V)$ and $v \in V$. Let $\varepsilon_{T,v} \colon F[X] \to V$ be the map of *evaluation at T and v*, given by

$$\varepsilon_{T,v}(f) = f(T)(v).$$

This map is a vector space homomorphism. Its image

$$\begin{aligned} Z(T;v) &= im(\varepsilon_{T,v}) = \{f(T)(v) \mid f \in F[X]\} \\ &= span\{v, T(v), \ldots, T^i(v), \ldots\} \end{aligned} \tag{4.75}$$

is called the *T-cyclic subspace of V generated by v*. Thus, $Z(T;v)$ is the unique smallest T-invariant subspace of V containing v. While $\varepsilon_{T,v}$ cannot be a ring homomorphism (as V is not a ring), note that for $f, g \in F[X]$,

$$\varepsilon_{T,v}(gf) = g(T)((\varepsilon_{T,v})(f)).$$

Hence, $ker(\varepsilon_{T,v})$ is an ideal of $F[X]$. Note that elements of $ker(\varepsilon_{T,v})$ correspond to linear dependence relations among $v, T(v), T^2(v), \ldots$. The unique monic generator of the nonzero ideal $ker(\varepsilon_{T,v})$ of $F[X]$ is called the *T-annihilator of v*, denoted $m_{T,v}$. Thus, for $f \in F[X]$,

$$m_{T,v} | f \text{ in } F[X] \text{ iff } f(T)(v) = 0. \tag{4.76}$$

The FHT gives a vector space isomorphism

$$Z(T;v) \cong F[X]/(m_{T,v}).$$

4.64. Take any $T \in \mathcal{L}(V)$ and any $v \in V$.

(i) Prove that

$$dim(Z(T;v)) = deg(m_{T,v}), \tag{4.77}$$

and if this degree is d, then $\mathcal{B} = \{v, T(v), T^2(v), \ldots, T^{d-1}(v)\}$ is a base of $Z(T;v)$.

(ii) Prove that in $M_d(F)$,

$$[T|_{Z(T;v)}]_{\mathcal{B}} = C_{m_{T,v}}, \tag{4.78}$$

where the $C_{m_{T,v}}$ is the companion matrix of $m_{T,v}$ as in (4.73).

4.65. Let $T \in \mathcal{L}(V)$, and suppose that V is T-*cyclic*, i.e., $V = Z(T;v)$ for some $v \in V$. Prove that

$$m_T = m_{T,v} = \chi_T.$$

(Use (4.78) and problem 4.63 for the second equality.)

4.66. Let $T \in \mathcal{L}(V)$, let $v \in V$, and let $g \in F[X]$. Prove that

$$m_{T,g(T)(v)} = m_T / gcd(g, m_{T,v})$$

Here, *gcd* means the unique monic *gcd*. (Hint: Recall problem 3.56.)

4.67. Let $T \in \mathcal{L}(V)$, and suppose that V is T-cyclic, say $V = Z(T;v)$.

(i) Let W be any T-invariant subspace of V, and let $\overline{T}$ be the linear transformation in $\mathcal{L}(V/W)$ induced by T. Note that $V/W = Z(\overline{T};\overline{v})$, where $\overline{v} = v + W$. Let $h = m_{\overline{T},\overline{v}}$, and let $w = h(T)(v)$. Prove that

$$W = Z(T;w).$$

(Hint: Apply (4.77) and problem 4.66.) Thus, every T-invariant subspace of a T-cyclic vector space is again T-cyclic.

(ii) Prove that there is a one-to-one correspondence between T-invariant subspaces of $Z(T;v)$ and monic divisors of $m_{T,v}$ in $F[X]$.

4.68. *Cayley–Hamilton Theorem.* Take any $T \in \mathcal{L}(V)$.

(i) Prove that $m_T | \chi_T$ in $F[X]$, or, equivalently, that $\chi_T(T) = 0$ in $\mathcal{L}(V)$. This is the Cayley–Hamilton Theorem. (Hint: If V is T-cyclic, apply problem 4.65. If not, let $W = Z(T;v)$ for some nonzero $v \in V$, and use (4.55) and problem 4.53(iii).)

(ii) Prove also that if q is irreducible in $F[X]$ and $q|\chi_T$, then $q|m_T$. Thus, the irreducible factors of χ_T are the same as the irreducible factors of m_T.

4.69. Let $T \in \mathcal{L}(V)$. Prove that there is $v \in V$ with $m_{T,v} = m_T$. (Prove this first when m_T is a power of an irreducible polynomial, then use the primary decomposition for the general case.)

4.70. Let $T \in \mathcal{L}(V)$. Prove that $\{v \in V \mid m_{T,v} \neq m_T\}$ is a finite union of proper subspaces of V. Thus, if $|F| = \infty$, then "nearly all" $v \in V$ satisfy $m_{T,v} = m_T$.

4.71. Let $T \in \mathcal{L}(V)$. Prove that if $m_T = \chi_T$, then V is T-cyclic.

If R is a ring and A is a subring of R, then the *centralizer of A in R* is

$$C_R(A) = \{r \in R \mid ra = ar \text{ for all } a \in A\}, \tag{4.79}$$

which is a subring of R.

4.72. Let $T \in \mathcal{L}(V)$, and suppose that V is T-cyclic. Prove that $C_{\mathcal{L}(V)}(F[T]) = F[T]$.

4.73. A linear transformation $T \in \mathcal{L}(V)$ is said to be *semisimple* if every T-invariant subspace W of V has a *T-invariant complement*, i.e., a T-invariant subspace Y of V such that $V = W \oplus Y$.

(i) Prove that if T is semisimple and W is any T-invariant subspace of V, then $T|_W$ is semisimple in $\mathcal{L}(W)$.

(ii) Suppose that $m_T = q^r$ where q is irreducible in $F[X]$. Prove that T is semisimple iff $r = 1$. (For "if", use problem 4.49.)

(iii) Prove that T is semisimple iff $m_T = q_1 q_2 \dots q_k$, where the q_i are *distinct* monic irreducibles in $F[X]$.

4.8. Projection maps

Throughout this section, let V be a finite-dimensional F-vector space.

Projection maps. A linear transformation $P \in \mathcal{L}(V)$ is called a *projection* if $P^2 = P$ (i.e., P is an idempotent of the ring $\mathcal{L}(V)$). Then,

$$V = \mathit{im}(P) \oplus \mathit{ker}(P),$$

with $P|_{\mathit{im}(P)} = \mathit{id}_{\mathit{im}(P)}$ and $P|_{\mathit{ker}(P)} = 0$. Thus, P is completely determined by the subspaces $\mathit{im}(P)$ and $\mathit{ker}(P)$ of V, which (when nontrivial) are the eigenspaces of P. This P is sometimes called the

projection of V onto $im(P)$ along $ker(P)$. Note that $id_V - P$ is also a projection, onto $ker(P)$ along $im(P)$.

If $V = W_1 \oplus \ldots \oplus W_k$ is any direct sum decomposition of V, then there is an associated family of projections $P_1, \ldots, P_k$ in $\mathcal{L}(V)$ defined by $P_i|_{W_i} = id_{W_i}$ and $P_i|_{W_j} = 0$ for $j \neq i$. Note that for all i, j,

$$P_i^2 = P_i, \quad P_i P_j = 0 \text{ if } i \neq j, \quad \text{and} \quad P_1 + \ldots + P_k = id_V\,. \tag{4.80}$$

Conversely, given $P_1, \ldots, P_k \in \mathcal{L}(V)$ satisfying the conditions in (4.80), we have

$$V = im(P_1) \oplus \ldots \oplus im(P_k),$$

and $P_1, P_2, \ldots, P_k$ is the associated family of projections.

4.74. Let $s_1, \ldots, s_k \in F[X]$ with $k \geq 2$. Suppose that $deg(s_i) \geq 1$ for each i and that $gcd(s_i, s_j) \sim 1$ in $F[X]$ whenever $i \neq j$. Let $f = s_1 s_2 \ldots s_k$ and let

$$g_i = f/s_i = s_1 s_2 \ldots s_{i-1} s_{i+1} \ldots s_k \qquad \text{for } i = 1, 2, \ldots, k.$$

Note that by induction on i, $gcd(g_1, \ldots, g_i) \sim s_{i+1} \ldots s_k$. Hence, $gcd(g_1, \ldots, g_k) \sim 1$, so as $F[X]$ is a PID, the ideal $(g_1, \ldots, g_k) = (1)$ in $F[X]$. Therefore, there exist $h_1, \ldots, h_k \in F[X]$ with

$$h_1 g_1 + \ldots + h_k g_k = 1.$$

Let $p_i = h_i g_i$, for each i, so that $p_1 + \ldots + p_k = 1$. Prove that for all i, j, we have $f \,|\, p_i p_j$ whenever $i \neq j$ and $f \,|\, p_i(1 - p_i)$.

4.75. Let $T \in \mathcal{L}(V)$ and let m_T have irreducible factorization $m_T = q_1^{r_1} \ldots q_k^{r_k}$ where the q_i are distinct monic irreducibles in $F[X]$. As in problem 4.56, let

$$U_i = ker\left(q_i^{r_i}(T)\right) = im\left(q_1^{r_1} \ldots q_{i-1}^{r_{i-1}} q_{i+1}^{r_{i+1}} \ldots q_k^{r_k}(T)\right)$$

for $i = 1, 2, \ldots, k$, so that

$$V = U_1 \oplus \ldots \oplus U_k$$

is the primary decomposition of V relative to T. In the notation of the preceding problem, let $s_i = q_i^{r_i}$, so $f = m_T$, $g_i = f/s_i$, and $p_i = h_i g_i$, where $h_1 g_1 + \ldots + h_k g_k = 1$ in $F[X]$.

(i) Let $P_i = p_i(T) \in F[T] \subseteq \mathcal{L}(V)$. Prove that $P_1, \ldots, P_k$ satisfy the conditions of (4.80).

(ii) Prove that $im(P_i) = U_i$, for each i. Deduce that $P_1, \ldots, P_k$ is the family of projections associated to the direct sum decomposition $V = U_1 \oplus \ldots \oplus U_k$.

4.76. Let $T \in \mathcal{L}(V)$ be a triangulable linear transformation with distinct eigenvalues $\lambda_1, \ldots \lambda_k$, minimal polynomial

$$m_T = (X - \lambda_1)^{r_1} \ldots (X - \lambda_k)^{r_k},$$

and generalized eigenspaces $U_i = ker\left((T - \lambda_i\, id_V)^{r_i}\right)$ as in (4.72). So, the primary decomposition of V relative to T is $V = U_1 \oplus \ldots \oplus U_k$. Let $P_1, \ldots, P_k$ be the family of projections associated to this direct sum decomposition of V. By the preceding problem, each $P_i \in F[T]$.

(i) Let

$$D = \lambda_1 P_1 + \ldots + \lambda_k P_k \in F[T]$$

and let $N = T - D$. Prove that $T = D + N$ with D diagonalizable, N nilpotent, and $DN = ND$.

(ii) Uniqueness of D and N: Suppose that $D', N' \in \mathcal{L}(V)$ and that $T = D' + N'$, with D' diagonalizable, N' nilpotent, and $D'N' = N'D'$. Prove that $D' = D$ and $N' = N$.

4.77. Suppose that $T \in \mathcal{L}(V)$ is diagonalizable, and let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. As noted in Example 4.57(i), the primary decomposition of V relative to T is then the eigenspace decomposition

$$V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k} \qquad \text{where} \quad V_{\lambda_i} = ker(T - \lambda_i\, id_V).$$

Problem 4.75 describes how we can find the family of projections associated with this direct sum decomposition of V. But we can do this more easily using the polynomials of Lagrange interpolation (see problem 4.11):

(i) Let

$$p_i = \frac{(X-\lambda_1)\ldots(X-\lambda_{i-1})(X-\lambda_{i+1})\ldots(X-\lambda_k)}{(\lambda_i-\lambda_1)\ldots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\ldots(\lambda_i-\lambda_k)} \in F[X],$$

for $i = 1, 2, \ldots, k$. Note that $p_i(\lambda_i) = 1$, while $p_i(\lambda_j) = 0$ whenever $j \neq i$. Prove that $p_1 + \ldots + p_k = 1$ in $F[X]$ (e.g., apply problem 4.11). Observe also that

$$m_T = (X - \lambda_1)(X - \lambda_2)\ldots(X - \lambda_k) \,\big|\, p_i p_j$$

whenever $j \neq i$.

(ii) Let $P_i = p_i(T) \in \mathcal{L}(V)$ for $i = 1, 2, \ldots, k$. Prove that $P_1, P_2, \ldots, P_k$ satisfy the conditions of (4.80), and that for all i, j,

$$P_i|_{V_{\lambda_i}} = id_{V_{\lambda_i}}, \quad \text{while} \quad P_i|_{V_{\lambda_j}} = 0 \text{ for } j \neq i.$$

Thus, $P_1, P_2, \ldots, P_k$ is the family of projections associated to the direct sum decomposition $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$.

(iii) Here is a significant way these projections can be used. Prove that

$$T = \lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_k P_k,$$

and more generally that for any $f \in F[X]$

$$f(T) = f(\lambda_1)P_1 + f(\lambda_2)P_2 + \ldots + f(\lambda_k)P_k. \tag{4.81}$$

If a matrix $A \in M_n(F)$ is diagonalizable, then calculations of powers of (or polynomials in) A are facilitated by the diagonalization: If $Q^{-1}AQ = D$ with $D = diag(d_1, \ldots, d_n)$, then $A^j = QD^jQ^{-1}$ with $D^j = diag(d_1^j, \ldots, d_n^j)$. Note that by using the projections onto the eigenspaces of A the calculations are still simpler: To apply (4.81) to compute A^j one only needs to know that A is diagonalizable, and to know its eigenvalues in order to compute the projection matrices P_i. One does not need to determine eigenvectors of A, nor a base change matrix Q, nor the inverse of Q. The next problem illustrates this.

4.78. Let $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$. Determine the projection matrices onto the eigenspaces of A as in problem 4.77 and use these to compute A^j for every j in $\mathbb{N}$.

4.9. Cyclic decomposition and rational and Jordan canonical forms

Throughout this section V is a finite-dimensional F-vctor space.

Take any $T \in \mathcal{L}(V)$. The *Cyclic Decomposition Theorem* says that V is a direct sum of T-cyclic subspaces,

$$V = Z(T; v_1) \oplus \ldots \oplus Z(T; v_m), \tag{4.82}$$

and that the v_i can be chosen so that for

$$f_i = m_{T,v_i} = \chi_{T|_{Z(T;v_i)}},$$

we have

$$f_2|f_1,\ f_3|f_2,\ \ldots,\ f_{i+1}|f_i,\ \ldots,\ f_m|f_{m-1}.$$

Moreover, subject to these divisibility conditions the f_i are uniquely determined by T; the f_i are called the *invariant factors* of the linear transformation T. (The v_i and the T-cyclic subspaces $Z(T;v_i)$ are not uniquely determined.) Note that

$$\begin{aligned} f_1 &= lcm(f_1,\ldots,f_m) \\ &= lcm\left(m_{T|_{Z(T;v_1)}},\ldots,m_{T|_{Z(T;v_m)}}\right) = m_T \end{aligned} \tag{4.83}$$

(see problem 4.54), and

$$\begin{aligned} &deg(f_1)+\ldots+deg(f_m) \\ &\quad = dim(Z(T;v_1))+\ldots+dim(Z(T;v_m)) = dim(V). \end{aligned} \tag{4.84}$$

The next few problems give a proof of the Cyclic Decomposition Theorem. The proof is similar to the that of the Fundamental Theorem for Finite Abelian Groups in problems 2.114 and 2.116. Indeed, each theorem is a special case of the structure theorem for finitely generated torsion modules over a PID.

4.79. Let $T \in \mathcal{L}(V)$, and let W be a T-invariant subspace of V, and let $\overline{T} \in \mathcal{L}(V/W)$ be the linear transformation induced by T. Suppose that $m_{\overline{T}} = \chi_{\overline{T}} = m_T$. By problem 4.71, V/W is $\overline{T}$-cyclic, say $V/W = Z(\overline{T}; v+W)$ for some $v \in V$; then $m_{\overline{T},v+W} = m_{\overline{T}} = m_T$. Prove that

$$V = W \oplus Z(T;v) \quad \text{and} \quad m_{T,v} = m_T.$$

4.80. *Existence of cyclic decomposition.* Let $T \in \mathcal{L}(V)$, and let $T^* \in \mathcal{L}(V)^*$ be its dual linear transformation.

(i) By problem 4.69 applied to T^* there is $y \in V^*$ with

$$m_{T^*,y} = m_{T^*} = m_T.$$

(See Example 4.48 for the second equality.) Let

$$Z = Z(T^*;y) \subseteq \mathcal{L}(V^*).$$

By problem 4.16(iv) there is a subspace W of V with $W^\perp = Z$. Moreover, W is T-invariant as Z is T^*-invariant (see problem 4.21(i)). For the induced linear transformation $\overline{T}$ in $\mathcal{L}(V/W)$, prove that

$$m_{\overline{T}} = \chi_{\overline{T}} = m_T.$$

(See problem 4.21(ii) and Example 4.48.)

(ii) It follows by the preceding problem that V/W is $\overline{T}$-cylic and that for any $v_1 \in V$ such that $V/W = Z(\overline{T}; v_1 + W)$,

$$V \,=\, Z(T; v_1) \oplus W \quad \text{and} \quad m_{T,v_1} = m_T.$$

Using this, prove by induction that there is a T-cyclic direct sum decomposition of V as in (4.82) with $m_{T,v_2} | m_{T,v_1}$, $\ldots$, $m_{T,v_i} | m_{T,v_{i-1}}$, $\ldots$, $m_{T,v_m} | m_{T,v_{m-1}}$. This proves the existence part of the Cyclic Decomposition Theorem.

4.81. *Uniqueness of invariant factors.* Let $T \in \mathcal{L}(V)$. For any T-invariant subspace W of V, and any $g \in F[X]$, we write

$$g(T) \cdot W \,= im(g(T)|_W) \,=\, \{g(T)(w) \mid w \in W\};$$

this is a T-invariant subspace of W. Note that for any $v \in V$, we have

$$g(T) \cdot Z(T; v) \,=\, Z(T; g(T)(v)).$$

(i) Take any $v \in V$ and any irreducible $q \in F[X]$. Prove that

$$\begin{cases} q(T) \cdot Z(T; v) = Z(T; v), & \text{if } q \nmid m_{T,v}; \\ dim\big(q(T) \cdot Z(T; v)\big) = dim(Z(T; v)) - deg(q), & \text{if } q | m_{T,v}. \end{cases}$$

(Recall problem 4.66.)

(ii) For $v \in V$ and q irreducible in $F[X]$ as in part (i) and any $\ell \in \mathbb{N}$, prove that

$$\begin{aligned} dim\big(q^{\ell-1}(T) \cdot Z(T; v)\big) - dim\big(q^{\ell}(T) \cdot Z(T; v)\big) \\ = \begin{cases} deg(q) & \text{if } q^{\ell} \mid m_{T,v}; \\ 0 & \text{if } q^{\ell} \nmid m_{T,v}. \end{cases} \end{aligned} \tag{4.85}$$

(iii) Let $V = Z(T; v_1) \oplus \ldots Z(T; v_m)$ be any T-cyclic direct sum decomposition of V, and set $g_i = m_{T,v_i}$ for all i. Let $q_1, q_2, \ldots, q_k$ be the distinct monic irreducible polynomials occurring in the irreducible factorizations of the g_j in $F[X]$. Then, each $g_j = q_1^{r_{1,j}} \ldots q_i^{r_{i,j}} \ldots q_k^{r_{k,j}}$ for nonnegative integers $r_{i,j}$. Prove that for any $\ell \in \mathbb{N}$, and for $i = 1, 2, \ldots, k$,

$$\begin{aligned} dim\big(q_i^{\ell-1}(T) \cdot V\big) \,-\, dim\big(q_i^{\ell}(T) \cdot V\big) \\ = \, deg(q_i) \cdot \big|\{j \mid r_{i,j} \geq \ell\}\big|. \end{aligned} \tag{4.86}$$

This formula shows that for all i, ℓ the number

$$s_{i,\ell} = \big|\{j \mid r_{i,j} \geq \ell\}\big|$$

is determined by T, independent of the choice of T-cyclic decomposition of V. Hence, T determines

$$\big|\{j \mid r_{i,j} = \ell\}\big| = s_{i,\ell} - s_{i,\ell+1}. \tag{4.87}$$

(iv) In the setting of part (iii), suppose further that for the $g_j = m_{T,v_j}$ we have $g_2|g_1, \ldots, g_{i+1}|g_i, \ldots, g_k|g_{k-1}$. Then, for the exponents $r_{i,j}$ as in part (iii) we have

$$r_{i,1} \geq r_{i,2} \geq \ldots \geq r_{i,m}.$$

Since the number of $r_{i,j}$ with a given value ℓ is uniquely determined by T, prove that the g_j are uniquely determined by T. This yields the uniqueness of the invariant factors of T.

Elementary divisors. Let $T \in \mathcal{L}(V)$ with invariant factors $f_1, \ldots, f_m$ in $F[X]$. Let $q_1, \ldots, q_k$ be the distinct irreducible monic divisors of f_1, and as in part (iii) above write

$$f_j = q_1^{r_{1,j}} \ldots q_i^{r_{i,j}} \ldots q_k^{r_{k,j}}.$$

The $q_i^{r_{i,j}}$ with $r_{i,j} \neq 0$ are called the *elementary divisors* of T. Thus, the invariant factors of T determine its elementary divisors, and vice versa. Note that if $V = Z(T; v_1) \oplus \ldots \oplus Z(T; v_m)$ with the m_{T,v_j} the invariant factors of T, then by taking the primary decomposition of each $Z(T; v_j)$ we obtain a T-cyclic direct sum decomposition of V where the minimal polynomials of T on the summands are the elementary divisors of T. Conversely, in any T-cyclic direct sum decomposition of V where the minimal polynomials of T on the summands are powers of irreducibles, these minimal polynomials are the elementary divisors of T. This follows from (4.86) and (4.87).

Rational canonical form. A matrix $B \in M_n(F)$ is said to be in *rational canonical form* if it has the block diagonal form

$$B = \begin{pmatrix} C_{f_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & C_{f_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & C_{f_m} \end{pmatrix}, \tag{4.88}$$

where the C_{f_i} are companion matrices of monic polynomials f_i as in (4.73), with $f_2 | f_1, \dots, f_{i+1} | f_i, \dots, f_m | f_{m-1}$. To simplify notation, we write

$$B = diag(C_{f_1}, \dots, C_{f_m}).$$

Note that if $T \in \mathcal{L}(V)$ has invariant factors $f_1, \dots, f_m$, then there is a base $\mathcal{B}$ of V such that $[T]_{\mathcal{B}} = B$. This $\mathcal{B}$ is built as a union of suitable bases of the direct summands in a T-cyclic decomposition of V (see (4.78)). The matrix B is called the *rational canonical form of* T, and B is uniquely determined by T, since the invariant factors of T are uniquely determined.

As pointed out in the comments above about elementary divisors, there is a T-cyclic direct sum decomposition of V for which the minimal polynomials of T on the summands are the elementary divisors of T. By building a base $\mathcal{B}'$ of V from bases of these summands, one can obtain a matrix $B' = [T]_{\mathcal{B}'}$ for T in block diagonal form where the diagonal blocks are companion matrices of the elementary divisors of T. Because the elementary divisors of T are uniquely determined, the matrix B' for T is uniquely determined except for the order of the diagonal blocks. However, the matrix B' is usually not in rational canonical form, since the associated polynomials do not satisfy the required divisibility condition.

4.82.

(i) Let $B = diag(C_{f_1}, \dots, C_{f_m})$ be a matrix in rational canonical form as in (4.88) (with each $f_{i+1} | f_i$). Prove that $f_1, \dots, f_m$ are the invariant factors of the multiplication-by-B linear transformation $L_B \in \mathcal{L}(F^n)$ as in (4.25).

(ii) Take any $A \in M_n(F)$. Prove that there is a unique matrix $B \in M_n(F)$ in rational canonical form as in (4.88) such that B is similar to A. *rational canonical form of* A. The f_i

are the invariant factors of $L_A \in \mathcal{L}(F^n)$ and are called the *invariant factors* of A. Thus, similarity to a unique matrix in rational canonical form is the matrix version of the Cyclic Decomposition Theorem.

(iii) The *elementary divisors* of $A \in M_n(F)$ are the elementary divisors of $L_A \in \mathcal{L}(F^n)$, i.e., the powers of irreducibles appearing in the factorizations of the invariant factors of A. Show that there is a matrix C similar to A such that C is in block diagonal form with diagonal blocks the companion matrices of the elementary divisors of A. Such a matrix C is uniquely determined by A except for the order of the diagonal blocks.

4.83. Let $A, B \in M_n(F)$.

(i) Prove that for $n \leq 3$, if $m_A = m_B$ and $\chi_A = \chi_B$, then A and B are similar.

(ii) Give an example of $A, B \in M_4(F)$ with $m_A = m_B$ and $\chi_A = \chi_B$ but A and B are not similar.

4.84. Let $T \in \mathcal{L}(V)$, and let $f_1, f_2, \ldots, f_m$ be the invariant factors of T. Prove that the centralizer $C_{\mathcal{L}(V)}(F[T])$ has dimension

$$\begin{aligned} &dim\left(C_{\mathcal{L}(V)}(F[T])\right) \\ &\quad = deg(f_1) + 3\,deg(f_2) + 5\,deg(f_3) + \ldots + (2m-1)\,deg(f_m). \end{aligned}$$

Thus, if V is not T-cyclic, i.e., $m \geq 2$, then as

$$dim(V) = deg(f_1) + \ldots + deg(f_m),$$

we have

$$dim(F[T]) = deg(m_T) = deg(f_1) < dim(V),$$

while

$$dim\left(C_{\mathcal{L}(V)}(F[T])\right) > dim(V).$$

4.85. Let F be a finite field, with $|F| = q$.

(i) Determine the number of nilpotent matrices in $M_2(F)$ and in $M_3(F)$.

(ii) Now determine the number of nilpotent matrices in $M_n(F)$ for any $n \in \mathbb{N}$.

4.86. Suppose that F is an infinite field and take $T \in \mathcal{L}(V)$. Prove that V is T-cyclic iff V has only finitely many T-invariant subspaces. (For "only if" use problem 4.67.)

4.87. Let $A \in M_n(F)$. Prove that A is similar to its transpose A^t. (Hint: Do this first for A the companion matrix of a monic polynomial.)

4.88. Suppose that F is a subfield of a field K. Let $A, B \in M_n(F)$. Prove that if A and B are similar in $M_n(K)$, then they are already similar in $M_n(F)$.

Jordan canonical form. Let $S \in \mathcal{L}(V)$, and suppose that V is S-cyclic with

$$m_S = \chi_S = (X - \lambda)^n$$

for some $\lambda \in F$. So, $dim(V) = deg(\chi_S) = n$. We know that $[S]_{\mathcal{B}} = C_{(X-\lambda)^n}$ for some base $\mathcal{B}$ of V. This companion matrix is not triangular, but S is triangulable by (4.60). We can obtain a triangular matrix for S as follows: Let $N = S - \lambda\, id_V \in \mathcal{L}(V)$. Then, $N^n = m_S(S) = 0$, but $N^{n-1} = (X - \lambda)^{n-1}(S) \neq 0$. Take any $v \in V$ with $N^{n-1}(v) \neq 0$. Since $N^n(v) = 0$, we have $m_{N,v} | X^n$ but $m_{N,v} \nmid X^{n-1}$. Hence, $m_{N,v} = X^n$, so $V = Z(N; v)$ by dimension count. Then,

$$\mathcal{B}' = \{N^{n-1}(v),\, N^{n-2}(v),\, \ldots,\, N^{n-i}(v),\, \ldots,\, N(v),\, v\}$$

is a base of V and the matrix $[N]_{\mathcal{B}'}$ has 1's on the first superdiagonal and 0's elsewhere. Since $S = \lambda\, id_V + N$, we have

$$[S]_{\mathcal{B}'} = J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda & 1 & \ldots & 0 & 0 \\ 0 & 0 & \lambda & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \lambda & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda \end{pmatrix} \in M_n(F). \qquad (4.89)$$

The matrix $J_{\lambda,n}$ is called an *elementary Jordan matrix.* (For $n = 1$, $J_{\lambda,1} = \big(\lambda\big) \in M_1(F)$.) The matrices $J_{\lambda,n}$ and $C_{(X-\lambda)^n}$ are similar in $M_n(F)$ since they are each matrices for S.

A matrix J of $M_n(F)$ is called a *Jordan matrix* if J is a block diagonal matrix with each diagonal block an elementary Jordan matrix.

Now take any $T \in \mathcal{L}(V)$ with T triangulable. Since every irreducible factor of m_T has degree 1 (see (4.60)), every elementary divisor of T has the form $(X-\lambda)^r$. Let $(X-\lambda_1)^{r_1}, \ldots, (X-\lambda_k)^{r_k}$ be the elementary divisors of T. (There can be repetitions among the λ_i and the r_i. The number of times a given $(X-\lambda)^r$ appears in the list of elementary divisors is the number of invariant factors f_j of T such that $(X-\lambda)^r | f_j$ but $(X-\lambda)^{r+1} \nmid f_j$.) As noted in preceding problem 4.82, there is a base $\mathcal{B}$ of V with

$$[T]_{\mathcal{B}} = diag\left(C_{(X-\lambda_1)^{r_1}}, \ldots, C_{(X-\lambda_k)^{r_k}}\right),$$

i.e., the matrix in block diagonal form with the i-th diagonal block the companion matrix $C_{(X-\lambda_i)^{r_i}}$. From the similarity of the matrices $C_{(X-\lambda_i)^{r_i}}$ and J_{λ_i,r_i}, it follows that there is a base $\mathcal{B}'$ of V such that $[T]_{\mathcal{B}'}$ is a Jordan matrix J in block diagonal form

$$J = diag\left(J_{\lambda_1,r_1}, \ldots, J_{\lambda_k,r_k}\right). \tag{4.90}$$

The matrix J is called "the" *Jordan canonical form* of T. Because the elementary divisors of T are uniquely determined, the Jordan form J of T is unique, except for the order of appearance of the diagonal blocks J_{λ_i,r_i}.

The matrix version of the Jordan canonical form says: For any $A \in M_n(F)$ with A triangulable there is a Jordan matrix J as in (4.90) such that A is similar to J. This J is called "the" Jordan canonical form of A, and is uniquely determined by A except for the order of the diagonal blocks J_{λ_i,r_i}.

4.89. Let $T \in \mathcal{L}(V)$ with T triangulable, and let λ be an eigenvalue of T. For $\ell \in \mathbb{N}$, let

$$t_{\lambda,\ell} = rk\left((T-\lambda\, id_V)^{\ell-1}\right) - rk\left((T-\lambda\, id_V)^{\ell}\right).$$

Prove that the number of times the elementary Jordan block $J_{\lambda,k}$ occurs in a Jordan form matrix J for T is $t_{\lambda,k} - t_{\lambda,k+1}$.

4.90. Suppose that $char(F) \neq 2$. Let $J = J_{\lambda,n}$ be the elementary Jordan matrix in $M_n(F)$ for the eigenvalue λ. Note that J^2 is a triangular matrix, so it has a Jordan form.

(i) Suppose that $\lambda \neq 0$. Determine the Jordan canonical form of J^2.

(ii) Suppose that $\lambda = 0$. Determine the Jordan canonical form of J^2. (There are two cases, depending on whether n is even or odd.)

4.91. Let A be an invertible matrix in $M_n(\mathbb{C})$. Since $\mathbb{C}$ is algebraically closed, A is triangulable, so it has a Jordan form.

(i) Prove that A has a "square root" in $M_n(\mathbb{C})$, i.e., a matrix $B \in M_n(\mathbb{C})$ such that $B^2 = A$.

(ii) Determine in terms of the Jordan form of A when there are only finitely many $B \in M_n(\mathbb{C})$ with $B^2 = A$. When there are only finitely many such B, determine how many.

4.92. *Real Jordan form.* Take any irreducible monic $f \in \mathbb{R}[X]$ with $deg(f) > 1$. Then, $deg(f) = 2$ (see problem 3.72), and f has no real roots but has two distinct complex conjugate roots in $\mathbb{C}$, say $a + ib$ and $a - ib$, where $a, b \in \mathbb{R}$ and $b \neq 0$. So,

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

(i) Prove that $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$ is similar to the companion matrix C_f. (This also holds for $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$.)

(ii) For $n \in \mathbb{N}$, let $R_{f,n}$ be the matrix in $M_{2n}(\mathbb{R})$ in block triangular form

$$R_{f,n} \;=\; \begin{pmatrix} \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & I_2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & I_2 & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & \ddots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & I_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \end{pmatrix}. \tag{4.91}$$

Every entry in the array is in $M_2(\mathbb{R})$, each diagonal matrix being the one of part (i), each matrix on the first superdiagonal is the identity matrix I_2 and all other entries are the zero matrix $\mathbf{0} \in M_2(\mathbb{R})$. (If $n = 1$, $R_{f,1}$ is the matrix of

part (i).) This $R_{f,n}$ is called a *real Jordan matrix*. Prove that $R_{f,n}$ is similar in $M_{2n}(\mathbb{R})$ to the companion matrix C_{f^n} of f^n.

(iii) Prove that every matrix A in $M_n(\mathbb{R})$ is similar to a matrix B in block diagonal form where each of the diagonal blocks is an elementary Jordan matrix as in (4.89) or a real Jordan matrix as in (4.91). Such a B is called a *real Jordan form* of A. Prove, moreover, that the matrix B is uniquely determined by A, except for the order of the diagonal blocks, and that in each $R_{f,n}$, b and $-b$ can be interchanged.

4.93. Let V be an n-dimensional vector space over $\mathbb{C}$, with a base $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$. Let $T \in \mathcal{L}(V)$ be the linear transformation defined by $T(v_i) = v_{i+1}$ for $i = 1, 2, \ldots, n-1$ and $T(v_n) = v_1$.

(i) Determine m_{T,v_1} and use this to determine m_T and χ_T.

(ii) Let

$$\omega = e^{2\pi i/n} = cos(\tfrac{2\pi}{n}) + i\, sin(\tfrac{2\pi}{n}) \in \mathbb{C}.$$

Prove that $1, \omega, \omega^2, \ldots, \omega^{n-1}$ are each eigenvalues of T, and for each of these eigenvalues find an eigenvector.

(iii) Since T has n different eigenvalues, it is diagonalizable. For the base $\mathcal{B}$ of V given above, determine the matrix $A = [T]_{\mathcal{B}} \in M_n(\mathbb{C})$. Then find an invertible matrix $P \in M_n(\mathbb{C})$ such that the matrix $D = P^{-1}AP$ is diagonal.

(iv) Now determine P^{-1} for your matrix P of part (c). (Hint: Consider the transpose of the equation $D = P^{-1}AP$.)

(v) A *circulant* matrix in $M_n(\mathbb{C})$ is a matrix of the form

$$C = \begin{pmatrix} c_1 & c_2 & c_3 & \ldots & c_n \\ c_n & c_1 & c_2 & \ldots & c_{n-1} \\ c_{n-1} & c_n & c_1 & \ldots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & c_4 & \ldots & c_1 \end{pmatrix}. \tag{4.92}$$

The ij-entry of C is c_ℓ where $\ell \equiv j - i + 1 \ (mod\ n)$. Note that

$$C = c_1 I_n + \sum_{i=1}^{n-1} c_{n+1-i} A^i$$

where A is the matrix of part (iii). Use this to determine $det(C)$.

4.94. Let p be a prime number, and let F be a field with $char(F) = p$. Let V be a p-dimensional F-vector space with base $\mathcal{B} = \{v_1, v_2, \ldots, v_p\}$ and, as in the preceding problem, let $T \in \mathcal{L}(V)$ be the linear transformation defined by $T(v_i) = v_{i+1}$ for $i = 1, 2, \ldots, p-1$ and $T(v_p) = v_1$.

(i) Prove that T is triangulable but not diagonalizable. Determine the Jordan canonical form of T and find a base $\mathcal{C}$ of V so that $[T]_{\mathcal{C}}$ is in Jordan form.

(ii) Let

$$C = \begin{pmatrix} c_1 & c_2 & c_3 & \ldots & c_p \\ c_p & c_1 & c_2 & \ldots & c_{p-1} \\ c_{p-1} & c_p & c_1 & \ldots & c_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & c_4 & \ldots & c_1 \end{pmatrix} \in M_p(F)$$

be a circulant matrix over F. Determine $det(C)$.

4.10. The exponential of a matrix

For any $A \in M_n(\mathbb{R})$, the *exponential of* A is defined to be

$$\begin{aligned} e^A &= I_n + A + \tfrac{1}{2}A^2 + \ldots + \tfrac{1}{i!}A^i + \ldots \\ &= \lim_{k\to\infty} \sum_{i=0}^{k} \tfrac{1}{i!}A^i \in M_n(\mathbb{R}). \end{aligned} \tag{4.93}$$

To clarify the meaning of this formula, let $B(0), B(1), \ldots, B(k), \ldots$ be an infinite sequence of matrices in $\mathbb{R}^{m\times n}$, with $B(k) = (b(k)_{ij})$. For $C = (c_{ij}) \in \mathbb{R}^{m\times n}$, we say that $\lim_{k\to\infty} B(k) = C$ if $\lim_{k\to\infty} b(k)_{ij} = c_{ij}$ for all i, j. Equivalently, if we set

$$\|C\| = \max_{1\le i\le m,\, 1\le j\le n} |c_{ij}| \in \mathbb{R}.$$

then $\lim_{k\to\infty} B(k) = C$ just when $\lim_{k\to\infty} \|C - B(k)\| = 0$. The infinite series $\sum_{i=0}^{\infty} B(k)$ is defined to be the limit of the sequence of partial sums $\lim_{k\to\infty} \sum_{i=0}^{k} B(i)$ (when the limit exists). To see that the infinite

series of matrices in (4.93) converges, first note that for $D, E \in M_n(\mathbb{R})$ and $r \in \mathbb{R}$,

$$\|D+E\| \le \|D\|+\|E\|, \ \|DE\| \le n\|D\|\|E\|, \text{ and } \|rD\| = |r|\,\|D\|.$$

Thus, for $A \in M_n(\mathbb{R})$, letting $B(m) = \frac{1}{m!}A^m$ for $m = 0, 1, \ldots$, we have for $m \ge 1$ and any i, j,

$$|b(m)_{ij}| \le \|B(m)\| \le \tfrac{1}{m!}n^{m-1}\|A\|^m \le \tfrac{1}{m!}(n\|A\|)^m.$$

Hence,

$$\sum_{m=0}^{k} |b(m)_{ij}| \le \sum_{m=0}^{k} \tfrac{1}{m!}(n\|A\|)^m \le e^{n\|A\|}.$$

Therefore, the sequence $\sum_{m=0}^{k} |b(m)_{ij}|$ for $k = 0, 1, 2, \ldots$ is bounded above as well as nondecreasing, so it converges. Hence, the infinite series $\sum_{m=0}^{\infty} b(m)_{ij}$ converges, since it converges absolutely. The limit is the ij-component of e^A.

Example 4.95. The following formulas for matrix exponentials are easy to prove:

(i) If $D = \mathit{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix in $M_n(\mathbb{R})$, then

$$e^D = \mathit{diag}\left(e^{\lambda_1}, \ldots, e^{\lambda_n}\right).$$

Likewise for matrices in block diagonal form.

(ii) For any $A, P \in M_n(\mathbb{R})$ with P invertible,

$$e^{PAP^{-1}} = Pe^AP^{-1}.$$

(iii) For any $A \in M_n(\mathbb{R})$,

$$e^{(A^t)} = (e^A)^t,$$

where ^{t} denotes the transpose.

4.96. Take any $A, B \in M_n(\mathbb{R})$. Prove that

$$\text{if } AB = BA, \text{ then } e^{A+B} = e^Ae^B.$$

(Note that since $AB = BA$, the binomial formula holds for $(A+B)^k$.) It follows, by taking $B = -A$, that e^A is invertible, with

$$(e^A)^{-1} = e^{-A}.$$

It is essential for the preceding problem that $AB = BA$. For example, let $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. Then,

$$e^A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \text{ and } e^B = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right), \text{ while } e^{A+B} = \frac{1}{2}\begin{pmatrix} e+e^{-1} & e-e^{-1} \\ e-e^{-1} & e+e^{-1} \end{pmatrix}.$$

4.97. Let $J = J_{\lambda,n}$ be an elementary Jordan matrix in $M_n(\mathbb{R})$ as in (4.89). Compute e^J. (Note that $J = \lambda I_n + N$ with $N \cdot \lambda I_n = \lambda I_n \cdot J$ and N nilpotent.)

4.98. Take any $a, b \in \mathbb{R}$.

(i) Let $A = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \in M_2(\mathbb{R})$. Prove that

$$e^A = \begin{pmatrix} e^a cos(b) & e^a sin(b) \\ -e^a sin(b) & e^a cos(b) \end{pmatrix}.$$

(ii) Let $R_{f,n}$ be the real Jordan matrix of (4.91). Compute $e^{R_{f,n}}$.

4.99. Take any $A \in M_n(\mathbb{R})$. Prove that

$$det(e^A) = e^{tr(A)}.$$

4.100. Let $A \in M_n(\mathbb{R})$ with $A^t = -A$. Such an A is called a *skew-symmetric* matrix. Prove that e^A is a *special orthogonal matrix*, i.e.,

$$(e^A)^t = (e^A)^{-1} \quad \text{and} \quad det(e^A) = 1.$$

4.101. Let $A \in M_n(\mathbb{R})$ be diagonalizable, with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Let $P_1, \ldots, P_k$ be the family of projections associated with the direct sum decomposition of $\mathbb{R}^n$ into the eigenspaces of A, as in problem 4.77.

(i) Prove that

$$e^A = e^{\lambda_1} P_1 + e^{\lambda_2} P_2 + \ldots + e^{\lambda_k} P_k.$$

(ii) Use this to compute e^A, where $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$.

Exponentials of matrices and solutions to linear differential equations. Let $\mathit{Diff}(\mathbb{R})$ denote the $\mathbb{R}$-vector space of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. For $Y = \big(y_{ij}(x)\big) \in \mathit{Diff}(\mathbb{R})^{m \times n}$ define $\frac{d}{dx}(Y)$ to be the $m \times n$ matrix of functions with ij-entry $\frac{d}{dx}(y_{ij}(x))$. Take $A \in M_n(\mathbb{R})$. One can show that the function e^{xA} (mapping $x \mapsto e^{xA}$ for all $x \in \mathbb{R}$) lies in $\mathit{Diff}(\mathbb{R})^{n \times n}$, and that

$$\tfrac{d}{dx}\big(e^{xA}\big) = Ae^{xA}.$$

Hence, if $\gamma_1(x), \ldots, \gamma_n(x)$ in $\mathit{Diff}(\mathbb{R})^{n\times 1}$ are the columns of e^{xA}, then each $\gamma_i(x)$ is a solution of the system of linear differential equations

$$\tfrac{d}{dx}(Y) \;=\; AY \quad \text{for} \quad Y \in \mathit{Diff}(\mathbb{R})^{n\times 1}. \tag{4.94}$$

It is known from the theory of differential equations that the solution space of the system (4.94) is n-dimensional. Hence, the solution space consists of linear combinations of the columns $\gamma_1(x), \ldots, \gamma_n(x)$. All the calculations of e^A in the preceding problems apply just as well for e^{xA}, and they yield explicit formulas for all the solutions of (4.94).

4.11. Symmetric and orthogonal matrices over $\mathbb{R}$

Inner product on $\mathbb{R}^n$. Fix $n \in \mathbb{N}$. The *inner product* (or scalar product, or dot product) on $\mathbb{R}^n$ is the function $B\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by: For $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ and $w = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ in $\mathbb{R}^n$,

$$B(v,w) \;=\; a_1b_1 + \ldots + a_ib_i + \ldots + a_nb_n. \tag{4.95}$$

In matrix terms, when we identify $\mathbb{R}$ with $M_1(\mathbb{R})$

$$B(v,w) \;=\; v^t w, \tag{4.96}$$

where $v^t \in \mathbb{R}^{1\times n}$ is the transpose of v. Note the following key properties of the inner product:

- B is *bilinear*, i.e., for all $c, d \in \mathbb{R}$ and $v, w, v_1, v_2, w_1, w_2 \in \mathbb{R}^n$,

$$B(cv_1 + dv_2, w) \;=\; cB(v_1,w) + dB(v_2,w)$$
$$\text{and} \quad B(v, cw_1 + dw_2) \;=\; cB(v,w_1) + dB(v,w_2).$$

- B is *symmetric*, i.e.,

$$B(v,w) \;=\; B(w,v) \quad \text{for all} \quad v, w \in \mathbb{R}^n.$$

- B is *positive definite*, i.e., for all $v \in \mathbb{R}^n$,

$$B(v,v) \;\geq\; 0, \quad \text{with} \quad B(v,v) = 0 \text{ iff } v = 0.$$

(For if $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, then $B(v,v) = a_1^2 + \ldots + a_n^2$.)

The basic notions of length and orthogonality in $\mathbb{R}^n$ are expressible in terms of the inner product: For $v \in \mathbb{R}^n$ the (Euclidean) *norm* (or length) of v is

$$\|v\| \;=\; \sqrt{B(v,v)}. \tag{4.97}$$

Vectors v, w in $\mathbb{R}^n$ are said to be *orthogonal* (perpendicular), written $v \perp w$, just when $B(v,w) = 0$.

4.102. Let V be an $\mathbb{R}$-subspace of $\mathbb{R}^n$, and let $V^* = \mathcal{L}(V, \mathbb{R})$ be its dual space as in §4.3. Note that there is a linear transformation

$$\psi\colon V \to V^* \quad \text{given by } \psi(v)(u) = B(v,u) \quad \text{for all} \quad v, u \in V.$$

Since $\psi(v)(v) = \|v\|^2$ which is nonzero if $v \neq 0$, the map ψ is injective, hence an isomorphism as $dim(V^*) = dim(V)$.

(i) Let W be any subspace of V, and $W^\perp$ the corresponding subspace of V^* as in (4.39) The *orthogonal complement of W in V* is

$$\begin{aligned} W^{\perp,V} \;&=\; \psi^{-1}(W^\perp) \\ &=\; \{v \in V \mid B(v,w) = 0 \text{ for all } w \in W\}. \end{aligned} \tag{4.98}$$

Note that

$$dim(W^{\perp,V}) \;=\; dim(W^\perp) \;=\; dim(V) - dim(W)$$

(see (4.40)). Prove that

$$V \;=\; W \oplus W^{\perp,V}.$$

(ii) Deduce that V has an *orthogonal base*, i.e., a base $\mathcal{B} = \{y_1, \ldots, y_k\}$ such that $B(y_i, y_j) = 0$ whenever $i \neq j$. Let $z_i = \frac{1}{\|y_i\|} y_i$ for each i. Then $\mathcal{C} = \{z_1, \ldots, z_k\}$ is an *orthonormal base* of V, i.e., $B(z_i, z_j) = 0$ whenever $i \neq j$ and $\|z_i\| = 1$ for each i.

4.103. *Gram-Schmidt orthogonalization.* Let V be a subspace of $\mathbb{R}^n$. Here is the algorithm for constructing an orthonormal base of V starting from any base called the *Gram-Schmidt orthogonalization process.* Let $\mathcal{B} = \{v_1, \ldots, v_k\}$ be any base of V. Let $y_1 = v_1 \neq 0$ and let $z_1 = \frac{1}{\|y_1\|} y_1$. For $i = 2, 3, \ldots, k$ define the y_i and z_i recursively by

$$y_i = v_i - \sum_{j=1}^{i-1} B(v_i, z_j) z_j \qquad \text{and} \qquad z_i = \tfrac{1}{\|y_i\|} y_i.$$

This is well-defined because $z_1, \ldots, z_{i-1} \in span\{v_1, \ldots, v_{i-1}\}$, while v_i is not in this span, so $y_i \neq 0$. Prove that $\{z_1, \ldots, z_k\}$ is an orthonormal base of V. (Geometrically, let $V_{i-1} = span\{v_1, \ldots, v_{i-1}\}$. Then, $\{z_1, \ldots, z_{i-1}\}$ is an orthonormal base of V_{i-1}, and the sum $\sum_{j=1}^{i-1} B(v_i, z_j)z_j$ is the orthogonal projection of v_i onto V_{i-1}, so $y_i \in V_{i-1}^{\perp,V}$.)

4.104. Let V be an $\mathbb{R}$-subspace of $\mathbb{R}^n$ with $dim(V) \geq 1$. A linear transformation $S \in \mathcal{L}(V)$ is said to be *self-adjoint* if

$$B(S(v), w) = B(v, S(w)) \quad \text{for all} \quad v, w \in V.$$

(i) Let $\mathcal{B} = \{z_1, \ldots, z_k\}$ be an orthonormal base of V. Prove that $S \in \mathcal{L}(V)$ is self-adjoint iff $[S]_{\mathcal{B}}$ is a symmetric matrix, i.e., $[S]_{\mathcal{B}} = ([S]_{\mathcal{B}})^t$.

(ii) If $S \in \mathcal{L}(V)$ is self-adjoint and W is any S-invariant subspace of V, prove that $W^{\perp,V}$ is also S-invariant.

(iii) Suppose that $S \in \mathcal{L}(V)$ is self-adjoint. We claim that V must have a 1-dimensional S-invariant subspace. For if not, V must have a 2-dimensional S-invariant subspace W, by problem 4.38. Then $S|_W \in \mathcal{L}(W)$ is self-adjoint, so that for any orthonormal base $\mathcal{C}$ of W, $[S|_W]_{\mathcal{C}}$ is a symmetric matrix in $M_2(\mathbb{R})$. By problem 4.45, $[S|_W]_{\mathcal{C}}$ is diagonalizable; hence, W contains a 1-dimensional S-invariant subspace spanned by an eigenvector. This proves the claim. Now prove that V has an orthonormal base consisting of eigenvectors of S. It follows in particular that S is diagonalizable.

Matrix multiplication in terms of the inner product. Take a matrix $A = [\alpha_1, \ldots, \alpha_m] \in \mathbb{R}^{n \times m}$, where each $\alpha_j \in \mathbb{R}^n$ is the j-th column of A. Likewise, let $C = [\gamma_1, \ldots, \gamma_k] \in \mathbb{R}^{n \times k}$. Observe that

$$A^t C = (b_{ij}) \in \mathbb{R}^{m \times k}, \quad \text{where each} \quad b_{ij} = B(\alpha_i, \gamma_j). \tag{4.99}$$

Orthogonal matrices and orthogonal similarity. A matrix Q in $M_n(\mathbb{R})$ is said to be an *orthogonal matrix* if

$$Q^t Q = I_n.$$

In view of (4.99), Q is orthogonal iff the columns of Q form an orthonormal base of $\mathbb{R}^n$. Suppose Q is orthogonal. Since

$$(Q^t)^t Q^t \;=\; QQ^t \;=\; I_n,$$

Q^t is also orthogonal. Hence, the rows of Q are an orthonormal base of $\mathbb{R}^{1\times n}$.

Two matrices $A, B \in M_n(\mathbb{R})$ are said to be *orthogonally similar* if there is an orthogonal matrix Q in $M_n(\mathbb{R})$ with $B = Q^{-1}AQ$. Orthogonal similarity is an equivalence relation on $M_n(\mathbb{R})$ since products and inverses of orthogonal matrices are orthogonal. It follows from problem 4.104(iii) that every symmetric matrix in $M_n(\mathbb{R})$ is orthogonally similar to a diagonal matrix. (The converse clearly holds: Every matrix in $M_n(\mathbb{R})$ orthogonally similar to a diagonal matrix is symmetric.)

4.105. Let V be a subspace of $\mathbb{R}^n$. A linear transformation $U \in \mathcal{L}(V)$ is called an *orthogonal transformation* if

$$B(Uv, Uw) \;=\; B(v, w) \quad \text{for all} \quad v, w \in V,$$

or, equivalently,

$$\|U(v)\| \;=\; \|v\|, \quad \text{for all} \quad v \in V.$$

These conditions are equivalent because $\|v\| = \sqrt{B(v,v)}$ and

$$B(v, w) \;=\; \tfrac{1}{2}\big(\|v+w\|^2 - \|v\|^2 - \|w\|^2\big).$$

Since U is distance-preserving, i.e, $\|U(v) - U(w)\| = \|v - w\|$ for all $v, w \in V$, this U is a rigid motion on V, as in problem 2.11. Note that the only possible eigenvalues of U in $\mathbb{R}$ are 1 and -1.

(i) Let $\mathcal{B}$ be an orthonormal base of V. Prove that $U \in \mathcal{L}(V)$ is an orthogonal transformation iff $[U]_{\mathcal{B}}$ is an orthogonal matrix.

(ii) Let $U \in \mathcal{L}(V)$ be an orthogonal transformation, and let W be a U-invariant subspace of V. Prove that $W^{\perp,V}$ is also U-invariant.

(iii) V is an *orthogonal sum* of subspaces $W_1, \ldots, W_k$ if $V = W_1 \oplus \ldots \oplus W_k$ and for any $w_i \in W_i$ and $w_j \in W_j$ with $i \neq j$, we have $w_i \perp w_j$. When this occurs, we write

$$V \;=\; W_1 \perp \ldots \perp W_k. \tag{4.100}$$

Let $U \in \mathcal{L}(V)$ be an orthogonal transformation. Prove that there are U-invariant subspaces $W_1, \ldots, W_k$ of V with $V = W_1 \perp \ldots \perp W_k$ such that for each i either

(a) $dim(W_i) = 1$ and $U|_{W_i} = id_{W_i}$; or

(b) $dim(W_i) = 1$ and $U|_{W_i} = -\,id_{W_i}$; or

(c) $dim(W_i) = 2$ and there is an orthonormal base $\mathcal{B}_i$ of W_i with $[T|_{W_i}]_{\mathcal{B}_i} = \begin{pmatrix} cos(\theta_i) & -sin(\theta_i) \\ sin(\theta_i) & cos(\theta_i) \end{pmatrix}$ for some $\theta_i \in \mathbb{R}$ with $sin(\theta_i) \neq 0$. (Recall problem 2.10.)

Thus, by taking a union of orthonormal bases of the W_i, we obtain an orthonormal base $\mathcal{B}$ of V such that $[U]_{\mathcal{B}} = A$ with A in block diagonal form

$$A = diag(B_1, \ldots, B_k), \tag{4.101}$$

where for each i, either

$$\begin{aligned} &B_i = (1) \in M_1(\mathbb{R}); \text{ or} \\ &B_i = (-1) \in M_1(\mathbb{R}); \text{ or} \\ &B_i = \begin{pmatrix} cos(\theta_i) & -sin(\theta_i) \\ sin(\theta_i) & cos(\theta_i) \end{pmatrix} \text{ with } sin(\theta_i) \neq 0. \end{aligned} \tag{4.102}$$

Note that A is a real Jordan form of U, and the diagonal blocks of A are thus uniquely determined by U, except that in any 2×2 block the terms $sin(\theta_i)$ and $-sin(\theta_i)$ can be interchanged. This gives a geometric interpretation of an orthogonal transformation U on V: The vector space V is an orthogonal sum of subspaces on which either U is the identity or minus the identity, or two-dimensional subspaces on each of which U acts by a rotation about the origin.

(iv) Deduce that a matrix $Q \in M_n(\mathbb{R})$ is orthogonal iff Q is orthogonally similar to a matrix A as in (4.101) and (4.102).

(v) Suppose that $dim(V) = 3$, and the orthogonal transformation $U \in \mathcal{L}(V)$ satisfies $det(U) = 1$. Prove that there is an orthonormal base $\mathcal{B} = \{z_1, z_2, z_3\}$ of V such that

$$[U]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{pmatrix},$$

for some $\theta \in \mathbb{R}$ with $\theta > 0$ (possibly $sin(\theta) = 0$). Thus, U can be viewed as a rotation through an angle θ about the axis through the origin determined by z_1.

The matrix A^tA. Take any $A \in \mathbb{R}^{m\times n}$. Then, A^tA is a symmetric matrix, so it is (orthogonally) diagonalizable in $M_n(\mathbb{R})$. Moreover, if μ is an eigenvalue of A^tA with eigenvector $v \in \mathbb{R}^n$, then

$$\|Av\|^2 \;=\; (Av)^t(Av) \;=\; v^t(A^tAv) \;=\; \mu\,\|v\|^2,$$

hence $\mu \geq 0$. If A^tA is similar to $diag(\mu_1, \ldots, \mu_n)$, (so the μ_i are the eigenvalues of A^tA) then the nonnegative real numbers $\sqrt{\mu_1}, \ldots, \sqrt{\mu_n}$ are called the *singular values* of A.

4.106. Let $A \in \mathbb{R}^{m\times n}$. Note that for $v \in \mathbb{R}^n$, if $A^tAv = 0$, then $\|Av\|^2 = v^t(A^tAv) = 0$, so $Av = 0$. Now prove the equality of column spaces,

$$Col(A^tA) \;=\; Col(A^t).$$

(Hint: Compare dimensions.)

4.107. *Best approximate solutions of systems of linear equations.* A system of m linear equations in n unknowns over $\mathbb{R}$ can be restated as a single matrix equation

$$A\,x \;=\; b \tag{4.103}$$

with $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$. Given A and b, the system is solved by finding $x \in \mathbb{R}^n$ for which equation (4.103) holds. Clearly, a solution exists iff $b \in Col(A)$. When there is no solution, one would often like the best approximate solution.

(i) Prove that there is always a solution $y \in \mathbb{R}^n$ to the system

$$A^tA\,y \;=\; A^tb.$$

(Apply the preceding problem.)

(ii) Prove that a solution y to $A^tAy = A^tb$ is the best possible approximate solution to $Ax = b$ by showing that

$$b - Ay \;\in\; Col(A)^{\perp,\mathbb{R}^m}.$$

Thus, Ay is the element of $Col(A)$ closest to b in $\mathbb{R}^m$. In particular, if $b \in Col(A)$, show that $Ay = b$.

4.108. *Determinant as volume.* Let A be an invertible matrix in $M_n(\mathbb{R})$. This problem gives a justification for the standard geometric interpretation of $|det(A)|$ as the volume of the parallelepiped in $\mathbb{R}^n$ determined by the columns of A.

(i) Prove that there is an orthonormal base $\mathcal{B} = \{z_1, \ldots, z_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of the symmetric matrix A^tA. (See problem 4.104(iii) and the comments preceding problem 4.105.)

(ii) For the eigenvectors z_i of part (i), say $A^tAz_i = \mu_i z_i$, with $\mu_i > 0$ in $\mathbb{R}$. (See the comments preceding problem 4.106. Note that $\mu_i \neq 0$ as A^tA is invertible.) Let $y_i = Az_i$, for each i. Prove that $\|y_i\| = \sqrt{\mu_i}$ and that $y_i \perp y_j$ whenever $i \neq j$, and that

$$|det(A)| = \sqrt{\mu_1}\sqrt{\mu_2}\ldots\sqrt{\mu_n}.$$

We can interpret this as follows: Let

$$\mathcal{U} = \{c_1z_1 + \ldots + c_nz_n \mid \text{each } c_i \in \mathbb{R} \text{ and } 0 \leq c_i \leq 1\},$$

which we can consider a unit hypercube in $\mathbb{R}^n$ with volume $Vol(\mathcal{U}) = 1$, since it has mutually perpendicular edges at any vertex with each edge of length 1. (The edges at the origin are the z_i.) Let

$$\mathcal{R} = A(\mathcal{U}) = \{Av \mid v \in \mathcal{U}\}.$$

Then,

$$\mathcal{R} = \{c_1y_1 + \ldots + c_ny_n \mid \text{each } c_i \in \mathbb{R} \text{ and } 0 \leq c_i \leq 1\},$$

which can be considered a hyperrectangle, since at any vertex the edges are mutually perpendicular. So, since the edges of $\mathcal{R}$ with vertex at the origin are the y_i,

$$Vol(\mathcal{R}) = \|y_1\|\ldots\|y_n\| = \sqrt{\mu_1}\ldots\sqrt{\mu_n} = |det(A)|.$$

Likewise, for any hyperrectangle $\mathcal{H}$ with edges parallel to the z_i, we have $A(\mathcal{H})$ is a hyperrectangle with edges parallel to the y_i, with i-th edge length $\sqrt{\mu_i}$ times the i-th edge length of $\mathcal{H}$; hence,

$$Vol(A(\mathcal{H})) = |det(A)|Vol(\mathcal{H}).$$

For any bounded solid $\mathcal{T}$ in $\mathbb{R}^n$ with a reasonable boundary, $\mathcal{T}$ is expressible as a union of (possible infinitely many) hyperrectangles each

with sides parallel to the z_i which intersect only along their boundaries. Then $Vol(\mathcal{T})$ is the sum of the volumes of these hyperrectagles, and

$$Vol(A(\mathcal{T})) \;=\; |det(A)|Vol(\mathcal{T}).$$

In particular, taking the standard unit hypercube $\mathcal{S}$ determined by the orthonormal standard base $\{\varepsilon_1, \ldots, \varepsilon_n\}$ of $\mathbb{R}^n$, then $\mathcal{P} = A(\mathcal{S})$ is the parallelepiped determined by the columns $A\varepsilon_1, \ldots, A\varepsilon_n$ of A, and

$$Vol(\mathcal{P}) \;=\; |det(A)|\; Vol(\mathcal{S}) \;=\; |det(A)|.$$

4.12. Group theory problems using linear algebra

4.109. For any prime $p \geq 3$, let G be a nonabelian group of order p^3 in which every nonidentity element has order p.

(i) Prove that

$$G \;\cong\; \left\langle a, b, c \;\middle|\; \begin{array}{l} a^p = b^p = c^p = 1, \\ ba = ab,\ ca = ac,\ cb = abc \end{array} \right\rangle.$$

Thus, $G \cong \mathcal{H}(\mathbb{Z}_p)$ (see problem 2.90). (Hint: Take any $a, b \in G$ with $a \in Z(G)$, $a \neq 1_G$, and $b \notin \langle a \rangle$. Let $N = \langle a, b \rangle$, a normal subgroup of G, which is an elementary abelian p-group with $Aut(N) \cong GL_2(\mathbb{Z}_p)$ (see problem 4.3). Consider the Jordan form of any element of $Aut(N)$ of order p.)

(ii) Determine $|Aut(G)|$.

4.110. Let F be a finite field with $|F| = q$. Let $p = char(F)$. Then $p \neq 0$ since the prime subring P_F of F is finite, and p is a prime number since $P_F \cong \mathbb{Z}_p$ and P_F is an integral domain. Note that F is a vector space over the field P_F, Hence $q = p^{[F:P_F]}$, where $[F{:}P_F] = dim_{P_F}(F) \in \mathbb{N}$. This problem determines all possible values of the order $|A|$ of elements A of the group $GL_2(F)$. Recall from (2.61) that $|GL_2(F)| = (q^2 - 1)(q^2 - q)$.

(i) Suppose that the degree-2 polynomial χ_A is reducible in $F[X]$. Prove that $|A| \big| p(q-1)$.

(ii) Determine how many conjugacy classes there are of elements of order $p(q-1)$ in $GL_2(F)$, and how many elements there are in each conjugacy class.

(iii) Prove that there exist irreducible monic polynomials of degree 2 in $F[X]$. (For example, count the number of monic polynomials of degree 2 in $F[X]$, and the number of those that are reducible.) Deduce that there is $A \in GL_2(F)$ with χ_A irreducible.

(iv) Take A with χ_A irreducible. Then, $m_A = \chi_A$, and hence $F[A] \cong F[X]/(m_A)$ is a field, with

$$\big|F[A]\big| = |F|^{dim_F(F[A])} = q^2.$$

Prove that $|A|\big|(q^2-1)$, and that there is $B \in F[A]^*$ with $|B| = q^2-1$. (Recall problem 3.32.) Now take any $C \in F[A]^*$ Prove that $|C| = q^2-1$ iff $m_C \big| \big(X^{q^2-1}-1\big)$ and $m_C \nmid (X^d-1)$ for every divisor d of q^2-1 with $d < q^2-1$.

(v) Determine how many conjugacy classes there are of elements of order q^2-1 in $GL_2(F)$, and how many elements there are in each conjugacy class. (Problem 4.72 can be helpful.)

4.111. Let F be a finite field with $|F| = q$, and let $p = char(F)$. As noted for the preceding problem, p is a prime number and q is a power of p. Assume that q is odd. Let $G = GL_2(F)$, a group of order $(q^2-1)(q^2-q) = (q-1)^2q(q+1)$ (see (2.61)). This problem determines the structure and the number of the Sylow subgroups of G. First observe that G has the following subgroups:

$$\begin{aligned} D &= \{\left(\begin{smallmatrix} a & 0 \\ 0 & c \end{smallmatrix}\right) \mid a, c \in F^*\}, \quad \text{with } |D| = (q-1)^2; \\ B &= \{\left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right) \mid a, c \in F^*,\ b \in F\}, \quad \text{with } |B| = q(q-1)^2; \\ U &= \{\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right) \mid b \in F\}, \quad \text{with } |U| = q. \end{aligned}$$

It is easy to check that B is the normalizer of U in G. Note that if s is an odd prime divisor of $|G|$ and $s \neq p$, then $s|(q-1)$ or $s|(q+1)$, but not both.

(i) Prove that every p-Sylow subgroup of G is elementary abelian of order q, and determine the number of p-Sylow subgroups of G.

(ii) Let s be a prime number dividing $q-1$. Prove that every s-Sylow subgroup of G is isomorphic to $C_m \times C_m$ for some

$m \in \mathbb{N}$, and determine the number of s-Sylow subgroups of G.

(iii) Now let s be a prime number dividing $q+1$. Prove that every s-Sylow subgroup of G is cyclic, and determine the number of s-Sylow subgroups of G.

(iv) Give a presentation by generators and relations of a 2-Sylow subgroup of G, and determine the number of 2-Sylow subgroups of G. Prove that the 2-Sylow subgroups are all non-abelian and nondihedral. There are two cases, depending on whether $q \equiv 1 \ (mod\ 4)$ or $q \equiv 3 \ (mod\ 4)$. (For the case $q \equiv 3 \ (mod\ 4)$, note that for $A \in G$, if m_A is irreducible of degree 2 in $F[X]$, then the roots of m_A in the field $F[A]$ are A and A^q.)

4.112. For elements b and c of finite order in a group G, we have seen in problem 2.23 that if $bc = cb$ then the group orders $|b|$ and $|c|$ give strong constraints on the possible values of $|bc|$. (In particular, $|bc| \,\big|\, lcm(|b|,|c|)$.) This problem shows that if we do not assume that $bc = cb$, then $|b|$ and $|c|$ give essentially no information about $|bc|$. We work in the group $GL_2(\mathbb{C})$. For $k \in \mathbb{Z} \setminus \{0\}$, let

$$\omega_k \;=\; e^{2\pi i/k} \;=\; cos(\tfrac{2\pi}{k}) + i\,sin(\tfrac{2\pi}{k}) \;\in\; \mathbb{C}^*.$$

Note that ω_k has order $|k|$ in $\mathbb{C}^*$ and that $\omega_k \cdot \omega_{-k} = 1$. Let

$$A_k \;=\; \begin{pmatrix} \omega_k & 0 \\ 0 & \omega_{-k} \end{pmatrix} \;\in\; GL_2(\mathbb{C}).$$

Note that

$$\chi_{A_k} \;=\; X^2 - 2cos(\tfrac{2\pi}{k})X + 1,$$

and that A_k has order $|k|$ in $GL_2(\mathbb{C})$. Take any $r, s \in \mathbb{N}$ with $r \geq 3$ and $s \geq 3$.

(i) Take any $t \in \mathbb{N}$ with $t \geq 3$. Prove that there is a matrix $P \in GL_2(\mathbb{C})$ such that

$$tr\big(A_r P A_s P^{-1}\big) \;=\; 2cos(\tfrac{2\pi}{t}).$$

Prove that $A_r P A_s P^{-1}$ is similar to A_t. Thus, for $B = A_r$ and $C = P A_s P^{-1}$, we have

$$|B| = r, \quad |C| = s, \quad \text{and} \quad |BC| = t.$$

(ii) Prove that there is a matrix $Q \in GL_2(C)$ such that

$$\left|tr(A_r Q A_s Q^{-1})\right| > 2.$$

Prove that $|A_r Q A_s Q^{-1}| = \infty$. Thus, for $B = A_r$ and $D = Q A_s Q^{-1}$, we have

$$|B| = r, \quad |D| = s, \quad \text{and} \quad |BD| = \infty.$$

Chapter 5

Fields and Galois Theory

This chapter gives problems on fields and their extensions, particularly finite degree extensions, for which Galois theory is an essential tool. Throughout the chapter, F is a field. We say that $F \subseteq K$ are fields if K is a field containing F as a subfield; then K is said to be an *extension field* of F. Since F is a field, its characteristic $char(F)$ (see (3.25) and p. 104) is either 0 or a prime number.

There is one change in notation from the previous chapter: If F is a field and $n \in \mathbb{N}$, then we set

$$F^n = \{c^n \mid c \in F\}. \tag{5.1}$$

Note that $F^n \setminus \{0\}$ is a subgroup of F^* with respect to multiplication. However, in general, F^n is not closed under addition, so not a ring. But there is an important exception: If $char(F) = p \neq 0$ (so p is a prime number), then F^{p^k} is closed under addition and subtraction (recall problem 3.22) for every $k \in \mathbb{N}$, and in fact F^{p^k} is a subfield of F.

5.1. Algebraic elements and algebraic field extensions

Let $F \subseteq K$ be fields. Then K is an F-vector space with the multiplication in K used for the scalar multiplication of F on K. The *degree of K over F* is defined to be

$$[K{:}F] \;=\; dim_F(K). \tag{5.2}$$

Thus, $[K{:}F] \in \mathbb{N}$ or $[K{:}F] = \infty$. Recall the easy but crucial *Tower Theorem*:

$$\text{if } F \subseteq L \subseteq K \text{ are fields, then } \;[K{:}F] \;=\; [L{:}F] \cdot [K{:}L]. \tag{5.3}$$

(Proof sketch: If $\{\alpha_i\}_{i\in I}$ is a base of L as an F-vector space and $\{\beta_j\}_{j\in J}$ is a base of K as an L-vector space, then $\{\alpha_i\beta_j\}_{i\in I,\, j\in J}$ is a base of K as an F-vector space.)

Let $F \subseteq K$ be fields, and fix $\alpha \in K$. Then define the *subring of K generated by α over F* to be

$$F[\alpha] \;=\; \{f(\alpha) \mid f \in F[X]\}. \tag{5.4}$$

This is clearly the smallest subring of K containing F and α, and it is an integral domain, since it is a subring of K. Similarly, define the *subfield of K generated by α over F* to be

$$\begin{aligned} F(\alpha) \;&=\; \{f(\alpha)g(\alpha)^{-1} \mid f, g \in F[X] \text{ and } g(\alpha) \neq 0\} \\ &=\; \text{the quotient field of } F[\alpha] \text{ in } K. \end{aligned} \tag{5.5}$$

Recall from (3.32) the evaluation homomorphism

$$\varepsilon_{F,\alpha}\colon F[X] \to K \quad \text{given by} \quad f \mapsto f(\alpha).$$

The map $\varepsilon_{F,\alpha}$ is a ring and F-vector space homomorphism. Clearly, $im(\varepsilon_{F,\alpha}) = F[\alpha] \subseteq K$ and $ker(\varepsilon_{F,\alpha})$ consists of 0 together with all the nonzero $f \in F[X]$ which have α as a root. The FHT yields

$$F[\alpha] \;\cong\; F[X]/\,ker(\varepsilon_{F,\alpha}), \tag{5.6}$$

a ring and F-vector space isomorphism. Since $F[\alpha]$ is an integral domain, $ker(\varepsilon_{F,\alpha})$ is a prime ideal of $F[X]$. There are two cases, depending on whether $ker(\varepsilon_{F,\alpha})$ is trivial:

Case I. α is said to be *transcendental over* F if $ker(\varepsilon_{F,\alpha}) = \{0\}$ in $F[X]$, i.e., α is not a root of any nonzero polynomial in F. When this occurs, $\varepsilon_{F,\alpha}^{-1}$ is a ring and F-vector space isomorphism,

$$F[\alpha] \cong F[X].$$

Moreover, $F(\alpha) \cong F(X)$ (the quotient field of $F[X]$); so

$$F(\alpha) = \{f(\alpha)g(\alpha)^{-1} \mid f, g \in F[X] \text{ and } g \neq 0\},$$

with

$$f_1(\alpha)g_1(\alpha)^{-1} = f_2(\alpha)g_2(\alpha)^{-1} \text{ iff } f_1g_2 = f_2g_1 \text{ in } F[X].$$

Note that $F[\alpha] \subsetneqq F(\alpha)$ and that $[F(\alpha):F] = \infty$, as

$$dim_F(F[\alpha]) = dim_F(F[X]) = \infty.$$

Case II. α is said to be *algebraic over* F if there is a nonzero $h \in F[X]$ with $h(\alpha) = 0$. That is, $ker(\varepsilon_{F,\alpha}) \supsetneqq \{0\}$. When this occurs, since $F[X]$ is a PID, the nonzero prime ideal $ker(\varepsilon_{F,\alpha})$ of $F[X]$ is a maximal ideal of $F[X]$ (recall problem 3.57(ii)). Hence, (5.6) shows that $F[\alpha]$ is a field, i.e.,

$$F(\alpha) = F[\alpha].$$

The ideal $ker(\varepsilon_{F,\alpha})$ is a principal ideal of the PID $F[X]$; the unique monic generator of this ideal is called the *minimal polynomial of* α *over* F, which is denoted $m_{F,\alpha}$.

Since $ker(\varepsilon_{F,\alpha})$ is a nonzero prime ideal of $F[X]$, the polynomial $m_{F,\alpha}$ is a prime element, hence irreducible in $F[X]$. As in problem 3.28, we have $\{1, \alpha, \alpha^2, \ldots, \alpha^{deg(m_{F,\alpha})-1}\}$ is a base of the F-vector space $F[\alpha]$; hence,

$$[F(\alpha):F] = dim_F(F[\alpha]) = deg(m_{F,\alpha}) < \infty. \tag{5.7}$$

From Case I and the Tower Theorem (5.3), it follows that if $[K:F] < \infty$ then every element α of K is algebraic over F, and $[F(\alpha):F] \big| [K:F]$.

Note 5.1. *Characterizations of the minimal polynomial.* For fields $F \subseteq K$, let $\alpha \in K$ with α algebraic over F. Take any monic $f \in F[X]$ with $f(\alpha) = 0$. Then, the following conditions are easily seen to be

equivalent:

(a) $f = m_{F,\alpha}$.

(b) For any $h \in F[X]$, if $h(\alpha) = 0$, then $f|h$ in $F[X]$.

(c) f is irreducible in $F[X]$.

(d) $deg(f) \leq [F(\alpha):F]$.

(e) $deg(f) = [F(\alpha):F]$.

If $F \subseteq L \subseteq K$ are fields, and $\alpha \in K$ is algebraic over F, then clearly α is also algebraic over L. But $m_{L,\alpha}$ need not be the same as $m_{F,\alpha}$. Indeed, $m_{L,\alpha}|m_{F,\alpha}$ in $L[X]$, (since $m_{F,\alpha} \in L[X]$ and $m_{F,\alpha}(\alpha) = 0$) but $m_{L,\alpha} = m_{F,\alpha}$ iff $m_{F,\alpha}$ is irreducible in $L[X]$ iff $m_{L,\alpha} \in F[X]$.

If $F \subseteq K$ are fields, and $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$, then we define recursively

$$\begin{aligned} F[\alpha_1, \alpha_2, \ldots, \alpha_n] &= F[\alpha_1, \alpha_2, \ldots, \alpha_{n-1}][\alpha_n], \\ F(\alpha_1, \alpha_2, \ldots, \alpha_n) &= F(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})(\alpha_n); \end{aligned} \tag{5.8}$$

so, $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the quotient field of $F[\alpha_1, \alpha_2, \ldots, \alpha_n]$ in K. Observe that $F[\alpha_1, \alpha_2, \ldots, \alpha_n] = F(\alpha_1, \alpha_2, \ldots \alpha_n)$ iff each α_i is algebraic over F, iff each α_i is algebraic over $F(\alpha_1, \ldots, \alpha_{i-1})$, iff $[F(\alpha_1, \alpha_2, \ldots, \alpha_n):F] < \infty$; when this occurs,

$$[F(\alpha_1, \alpha_2, \ldots, \alpha_n):F] \leq [F(\alpha_1):F] \cdot [F(\alpha_2):F] \cdot \ldots \cdot [F(\alpha_n):F] \tag{5.9}$$

If $S = \{\alpha_i\}_{i \in I}$ is an infinite subset of K, then define

$$F(S) = \bigcup_{n=1}^{\infty} \bigcup_{i_1, \ldots, i_n \in I} F(\alpha_{i_1}, \ldots, \alpha_{i_n}), \tag{5.10}$$

which is the subfield of K generated by S over F.

Recall *Kronecker's Theorem*: Let $f \in F[X]$ with $deg(f) \geq 1$. There is a field $K \supseteq F$ such that f has a root in K.

(Proof sketch: Let g be a monic irreducible factor of f in $F[X]$, let Y be a new indeterminate (different from X) and let $K' = F[Y]/(g(Y))$, which is a field. Then, $\widetilde{F} = \{c + (g(Y)) \mid c \in F\}$ is a subfield of K' with $F \cong \widetilde{F}$. Let $\widetilde{f}$ (resp. $\widetilde{g}$) be the polynomial in $\widetilde{F}[X]$ corresponding to f (resp. g) in $F[X]$. Then, $\overline{Y} = Y + (g(Y))$ is a root of $\widetilde{g}$, hence of $\widetilde{f}$ in K'. Take a set $K \supseteq F$ such that $|K \setminus F| = |K' \setminus \widetilde{F}|$, and

use a bijection $K' \setminus \widetilde{F} \to K \setminus F$ to define a field structure on K so that $K \cong K'$ by an isomorphism extending the given isomorphism $F \to \widetilde{F}$. Then f has a root in K since $\widetilde{f}$ has a root in K'.)

Note 5.2. *Splitting of polynomials.* Take any polynomial $f \in F[X]$ with $deg(f) \geq 1$. We say that f *splits over* a field $K \supseteq F$ if every irreducible factor of f in $K[X]$ has degree 1, i.e.,

$$f = c(X - \alpha_1)(X - \alpha_2)\dots(X - \alpha_k) \text{ for } c, \alpha_1, \alpha_2, \dots, \alpha_k \in K.$$

Note that when this occurs, $\alpha_1, \alpha_1, \dots, \alpha_k$ are roots of f in K. Moreover, very importantly, the α_i are the only roots of f in K or in any larger field containing K as a subfield. It follows from Kronecker's Theorem there is a field K containing F over which f splits. (For, by Kronecker, there is a field $L \supseteq F$ in which f has a root α_1. Then, $f = (X - \alpha_1)g$ in $L[X]$. By induction on degree, there is a field $K \supseteq L$ over which g splits; then f also splits over K.)

5.3. Let $F \subseteq K$ be fields. Let

$$A = \{\alpha \in K \mid \alpha \text{ is algebraic over } F\}.$$

This A is called the *algebraic closure of F in K*.

(i) Prove that A is a field with $F \subseteq A$.

(ii) Prove that if $\beta \in K$ is algebraic over A, then $\beta \in A$.

Algebraically closed fields. A field C is said to be *algebraically closed* if it satisfies the following equivalent conditions:

(a) C is algebraically closed in any field containing C.

(b) There is no field $K \supseteq F$ with $1 < [K{:}C] < \infty$.

(c) Every nonconstant polynomial in $C[X]$ has a root in C.

(d) Every irreducible polynomial in $C[X]$ has degree 1.

For example, the Fundamental Theorem of Algebra says that the field $\mathbb{C}$ of complex numbers is algebraically closed (see problem 5.101 below for a proof). Moreover, if F is any field, then F has an *algebraic closure*, i.e., an algebraically closed field $A \supseteq F$ such that A is algebraic over F (see problem 5.4 below). Such an A is actually unique up to isomorphism (see Note 5.56 below). If F is a subfield of $\mathbb{C}$, then the field $A_F = \{\alpha \in \mathbb{C} \mid \alpha \text{ is algebraic over } F\}$ is an algebraic

closure of F by problem 5.3. For example, $A_{\mathbb{Q}}$, an algebraic closure of $\mathbb{Q}$, is a subfield of $\mathbb{C}$, but $A_{\mathbb{Q}} \neq \mathbb{C}$ since $A_{\mathbb{Q}}$ is countable while $\mathbb{C}$ is uncountable.

5.4. *Existence of algebraic closures.* Let F be a field. Let S be an uncountable set containing F with $|S| > |F|$. Let

$$\mathcal{S} = \left\{ (K, +, \cdot) \;\middle|\; \begin{array}{l} F \subseteq K \subseteq S \text{ and } (K, +, \cdot) \text{ is an} \\ \text{algebraic extension field of } F \end{array} \right\}.$$

Here, $(K, +, \cdot)$ means the set K with specified operations of addition and multiplication making it a field. $\mathcal{S}$ is partially ordered by: $(K, +, \cdot) \leq (K', +, \cdot)$ just when K with its specified operations is a subfield of K' with its specified operations. Apply Zorn's Lemma to prove that $\mathcal{S}$ has a maximal element A. Prove that A is an algebraic closure of F.

5.5. Let $F \subseteq K$ be fields with $char(F) \neq 2$ and $[K{:}F] = 2$.

(i) Prove that there is $\beta \in K^*$ such that $\beta^2 \in F$ but $\beta \notin F$. Prove also that $K = F[\beta] = F(\beta)$. If $c = \beta^2$, we write $K = F(\sqrt{c}\,)$, though the expression $\sqrt{c}$ is ambiguous, as β and $-\beta$ are different square roots of c in K. But either choice yields the same field K. (There is in general no convention for choosing a preferred square root of an element of a field to call "the" square root of that element. The exception is that for $c \in \mathbb{R}$ with $c > 0$, its square root $\sqrt{c}$ customarily means the *positive* square root of c in $\mathbb{R}$.)

(ii) Prove that for $K = F(\sqrt{c}\,)$ as above,

$$K^2 \cap F \;=\; F^2 \cup cF^2.$$

Hence, if we also have $K = F(\sqrt{d}\,)$ for $d \in F$, then $dc^{-1} \in F^2$.

5.6. This problem shows the need for the assumption on $char(F)$ in the preceding problem. Let $F = \mathbb{Z}_2$, let $f = X^2 + X + 1 \in F[X]$ and let $K = F(\alpha)$, where $f(\alpha) = 0$. (Such a K and α exist by Kronecker's Theorem.)

(i) Prove that $m_{F,\alpha} = f$, and deduce that $[K{:}F] = 2$.

(ii) Prove that if $\gamma \in K$ and $\gamma^2 \in F$, then $\gamma \in F$. Hence, K is not obtainable by adjoining to F a square root of an element in F.

See problem 5.81 below for a characterization of quadratic extensions of a field of characteristic 2.

5.7. Let $F \subseteq K$ be fields with $char(F) \neq 2$ and $K = F\big(\sqrt{a}\,, \sqrt{b}\,\big)$ for some $a, b \in F$.

(i) Prove that $[K{:}F] = 4$ iff $a, b, ab \in F \setminus F^2$.

(ii) Assume that $[K{:}F] = 4$. Determine all the different fields L with $F \subseteq L \subseteq K$.

(iii) If $[K{:}F] = 4$, prove that $K = F\big(\sqrt{a} + \sqrt{b}\,\big)$.

5.8. Take any $r \in \mathbb{Q}$.

(i) Prove that $cos(r\pi)$ and $sin(r\pi)$ are algebraic over $\mathbb{Q}$. (Recall Euler's identity: $e^{i\theta} = cos(\theta) + i\, sin(\theta)$ in $\mathbb{C}$, where $i^2 = -1$.)

(ii) Suppose that $r = k/n$ with $k \in \mathbb{Z}$ and $n \in \mathbb{N}$. Prove that

$$[\mathbb{Q}(cos(r\pi)){:}\mathbb{Q}] \,\leq\, 2n \quad \text{and} \quad [\mathbb{Q}(sin(r\pi)){:}\mathbb{Q}] \,\leq\, 4n.$$

(See problem 5.122 below for the exact values of these degrees.)

5.9. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. Prove that

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

5.10. Let $F \subseteq K$ be fields, and let $\alpha \in K$ with α algebraic over F. Let $m_{F,\alpha} = X^n + c_{n-1}X^{n-1} + \ldots + c_0$. If $\alpha \neq 0$, prove that α^{-1} is algebraic over F, and determine $m_{F,\alpha^{-1}}$.

5.11. Let $F \subseteq K$ be fields, and let $\alpha \in K$ be algebraic over F. Take any $n \in \mathbb{N}$. Then, $F(\alpha^n)$ is a subfield of $F(\alpha)$.

(i) Prove that $[F(\alpha){:}F(\alpha^n)] \leq n$.

(ii) Prove that $[F(\alpha){:}F(\alpha^n)] = n$ iff $m_{F,\alpha} \in F[X^n]$.

(iii) Let $\beta \in K$ with β transcendental over F. Prove that $[F(\beta){:}F(\beta^n)] = n$.

5.12. Let $F \subseteq K$ be fields, and take $\alpha, \beta \in K$ with α and β algebraic over F. Prove that the following conditions are equivalent:

(a) $m_{F,\alpha}$ is irreducible in $F(\beta)[X]$

(b) $[F(\alpha,\beta):F] = [F(\alpha):F] \cdot [F(\beta):F]$.

(c) $m_{F,\beta}$ is irreducible in $F(\alpha)[X]$.

Compositum of fields. Let L and K be subfields of a field M. The *compositum of L and K*, denoted $L{\cdot}K$, is the subfield of M generated by L and K. That is, $L{\cdot}K$ is the quotient field in M of the ring $\big\{ \sum_{i=1}^{n} \alpha_i\beta_i \mid n \in \mathbb{N} \text{ and each } \alpha_i \in L \text{ and } \beta_i \in K \big\}$.

5.13. *Linear disjointness.* Let L and K be subfields of a field M, each of which has subfield F. Suppose that $[L:F] < \infty$.

(i) Prove that

$$[L{\cdot}K : K] \;\le\; [L : L \cap K] \;\le\; [L:F].$$

(ii) Prove that the following conditions are equivalent:

(a) Any $\alpha_1, \dots, \alpha_n$ in L that are linearly independent over F are also linearly independent over K.

(b) $[L{\cdot}K : K] = [L:F]$.

When the equivalent conditions (a) and (b) hold, we say that L and K are *linearly disjoint over F*.

Example 5.14. Take $\sqrt[3]{2} \in \mathbb{R}$, and let

$$\omega \;=\; e^{2\pi i/3} \;=\; \tfrac{1}{2}(-1 + i\sqrt{3}\,) \;\in\; \mathbb{C}^*,$$

so $\omega^3 = 1$ and $\omega^2 + \omega + 1 = 0$. Let

$$L \;=\; \mathbb{Q}(\sqrt[3]{2}\,) \qquad \text{and} \qquad K \;=\; \mathbb{Q}(\omega\sqrt[3]{2}\,).$$

Note that $X^3 - 2$ is irreducible in $\mathbb{Q}[X]$ since it has no root in $\mathbb{Q}$. Hence (see Note 5.1), $m_{\mathbb{Q},\sqrt[3]{2}} = X^3 - 2 = m_{\mathbb{Q},\omega\sqrt[3]{2}}$; thus,

$$K \;\cong\; \mathbb{Q}[X]/(X^3 - 2) \;\cong\; L \quad \text{and} \quad [K:\mathbb{Q}] \;=\; [L:\mathbb{Q}] \;=\; 3$$

(see (5.6)). Note that $\omega \notin K$ as $\omega \notin \mathbb{Q}$ and $2 = [\mathbb{Q}(\omega):\mathbb{Q}] \nmid [K:\mathbb{Q}]$. We have $L{\cdot}K = \mathbb{Q}(\sqrt[3]{2}\,,\omega) = K(\omega)$, so

$$[L{\cdot}K : K] \;=\; 2 \;<\; 3 \;=\; [L:\mathbb{Q}].$$

This shows that L and K are not linearly disjoint over $\mathbb{Q}$, even though $L \cap K = \mathbb{Q}$. But, $L \cdot K \ncong L \cdot L$, since $[L \cdot K : \mathbb{Q}] = 6$, while $L \cdot L = L$ and $[L{:}\mathbb{Q}] = 3$. This illustrates that the compositum of fields depends not just on the fields up to isomorphism, but also how they sit with respect to each other in the larger field.

5.15. Let f and g be irreducible monic polynomials in $F[X]$, and let K be a field containing F over which f and g split (which exists by repeated application of Kronecker's Theorem; see Note 5.2). Let

$$L = F(\alpha) \subseteq K, \quad \text{where } \alpha \text{ is a root of } f \text{ in } K,$$

and let

$$M = F(\beta) \subseteq K, \quad \text{where } \beta \text{ is a root of } g \text{ in } K.$$

Let $f = h_1^{r_1} \dots h_n^{r_n}$ be the monic irreducible factorization of f in $M[X]$ (i.e., the h_i are distinct monic irreducibles in $M[X]$, and each $r_i \in \mathbb{N}$), and let $g = k_1^{s_1} \dots k_m^{s_m}$ be the monic irreducible factorization of g in $L[X]$. Let γ_i be a root of h_i in K and δ_j a root of k_j, for all i, j. Prove that $m = n$ and that after rearranging the order of the k_i if necessary, for each i,

$$s_i = r_i \quad \text{and} \quad L(\delta_i) \cong M(\gamma_i),$$

hence,

$$deg(k_i)/\, deg(g) = deg(h_i)/\, deg(f).$$

(Hint: Consider $F[X,Y]/(f(X), g(Y))$.) Note that problem 5.12 above is the special case $n = 1$, $r_1 = 1$ of this problem.

5.16. Let $p_1, p_2, \dots, p_n$ be distinct prime numbers. Prove that

$$[\mathbb{Q}\big(\sqrt{p_1}\,, \sqrt{p_2}\,, \dots, \sqrt{p_n}\,\big) : \mathbb{Q}] = 2^n.$$

5.17. Prove that $\mathbb{Q}(\sqrt[3]{5}, \sqrt[3]{7}\,) = 9$.

5.2. Constructibility by compass and straightedge

One of the wonderful applications of the theory of fields is to the analysis of geometric constructions carried out using only a compass and straightedge. Field arguments allow one to show rather easily that certain desired constructions are actually impossible, thereby settling questions that had been open for millenia before being settled

in the nineteenth century. See any text covering fields for a discussion of this topic. One can start by defining a constructible number as a real number α such that $|\alpha|$ is the distance between two points in a compass-and-straightedge geometric construction. Then geometric arguments show that the set of constructible numbers is a subfield of $\mathbb{R}$ that is closed under taking square roots of positive elements. However, to focus the algebraic aspects of the theory, we will use the following definition of constructible numbers (which is equivalent to the geometric definition): An element $\alpha \in \mathbb{R}$ is a *constructible number* if there is a chain of fields

$$\mathbb{Q} = L_0 \subseteq L_1 \subseteq \ldots \subseteq L_k \subseteq \mathbb{R}$$

such that each $L_i = L_{i-1}(\sqrt{c_i}\,)$ for some $c_i \in L_{i-1}$ with $c_i \geq 0$, and $\alpha \in L_k$. The problems below are intended to be solved working from this definition, without recourse to geometric arguments. Let

$$E = \{\alpha \in \mathbb{R} \mid \alpha \text{ is constructible}\}. \tag{5.11}$$

It is clear from the definition of constructible numbers that if $\alpha \in E$ and $\alpha \geq 0$, then $\sqrt{\alpha} \in E$.

5.18.

(i) Prove that if α is a constructible number, then α is algebraic over $\mathbb{Q}$ and $[\mathbb{Q}(\alpha):\mathbb{Q}]$ is a power of 2.

(ii) Prove that the E of (5.11) is a subfield of $\mathbb{R}$ with $[E:\mathbb{Q}] = \infty$.

5.19. For the field E of constructible numbers, let M be a subfield of $\mathbb{C}$ with $M \supseteq E$ and $[M:E] = 2$.

(i) Prove that $M = E(\sqrt{-1}\,)$.

(ii) Prove that M is *quadratically closed*, i.e., there is no field $K \supseteq M$ with $[K:M] = 2$. (Equivalently, $M = M^2$.)

Constructible angles. All angles are measured in radians. An angle $\theta \in \mathbb{R}$ is said to be *constructible* if $\cos(\theta)$ is a constructible number. (θ is constructible iff θ (or $\theta' = \theta - k\pi$ with $k \in \mathbb{Z}$ and $0 \leq \theta' < \pi$) is an angle of a triangle constructed by compass and straightedge.) Note that $\cos(\theta)$ is a constructible number iff $\sin(\theta)$ is a constructible number.

5.20. Prove that

$$\mathbb{Q}(cos(\tfrac{\pi}{12})) = \mathbb{Q}(sin(\tfrac{\pi}{12})) = \mathbb{Q}(\sqrt{2}, \sqrt{3}).$$

It follows that $\frac{\pi}{12}$ is a constructible angle.

5.21.

(i) Prove that if θ and ψ are constructible angles then $\theta + \psi$ and $\theta - \psi$ are constructible angles.

(ii) Prove that if θ is constructible angle, then $\theta/2$ is a constructible angle.

Constructible polygons. For an integer $n \geq 3$, a regular n-gon (= regular polygon with n sides) is constructible (by compass and straightedge) iff $\frac{2\pi}{n}$ is a constructible angle—since $\frac{2\pi}{n}$ is the central angle of a regular n-gon. Note that problem 5.21(ii) shows that if a regular n-gon is constructible, then a regular $2n$-gon is also constructible.

5.22. Let $m, n \in \mathbb{N}$ with $m \geq 3$ and $n \geq 3$ and $gcd(m, n) = 1$. Prove that if a regular m-gon and a regular n-gon are constructible, then a regular mn-gon is constructible.

For an integer $n \geq 3$, let $\theta_n = \frac{2\pi}{n}$, and let $\gamma_n = cos(\theta_n)$. To assist in determining whether θ_n is a constructible angle, one would like to find polynomials in $\mathbb{Q}[X]$ with root γ_n. This is facilitated by considering

$$\omega_n = e^{2\pi i/n} = cos(\tfrac{2\pi}{n}) + i\,sin(\tfrac{2\pi}{n}) \in \mathbb{C}.$$

Note that ω_n is a primitive n-th root of unity in $\mathbb{C}$, i.e., an element of order n in the multiplicative group $\mathbb{C}^*$, hence a root of $X^n - 1 \in \mathbb{Z}[X]$. Note also that $\gamma_n = \frac{1}{2}(\omega_n + \omega_n^{-1})$. Suppose $f = c_k X^k + \ldots + c_0 \in \mathbb{Q}[X]$ has ω_n as a root and $k = deg(f)$ is even, and $c_i = c_{k-i}$ for all i, then the equation $\omega_n^{-k/2} f(\omega_n) = 0$ yields a polynomial with root γ_n.

5.23. Prove that a regular 3-gon (equilateral triangle) and 5-gon (pentagon) are constructible, while a regular 7-gon (heptagon) and 9-gon (nonagon) are not constructible. (Hint: ω_n is a root of $(X^n - 1)/(X - 1) \in \mathbb{Z}[X]$, and ω_9 is a root of $(X^9 - 1)/(X^3 - 1)$. Also a polynomial g of degree 3 in $\mathbb{Q}[X]$ is irreducible iff it has no

roots in $\mathbb{Q}$. For $g \in \mathbb{Z}[X]$, you can check for roots of g in $\mathbb{Q}$ by the Rational Roots Test; see problem 3.67.)

For more complete information on which regular n-gons are constructible and what elements of $\mathbb{R}$ are constructible, see problems 5.123 and 5.149 below.

5.3. Transcendental extensions

5.24. Let $F \subseteq K$ be fields, and let $s, t \in K$ with s and t each transcendental over F. Prove that if s is algebraic over $F(t)$ then t is algebraic over $F(s)$.

5.25. Let $F \subseteq K$ be fields, and let $t \in K$ with t transcendental over F. Take any $s \in F(t) \setminus F$. Thus, $s = f(t)/g(t)$ for some $f, g \in F[X]$ with $g \neq 0$. Choose f and g so that $gcd(f, g) \sim 1$.

(i) Prove that t is algebraic over $F(s)$.

(ii) Deduce that s is transcendental over F.

(iii) Prove that $sg - f$ is irreducible in the UFD $F[s][X]$.

(iv) Prove that $[F(t){:}F(s)] = max\big(deg(f), deg(g)\big)$.

It follows that $F(s) = F(t)$ iff $s = \frac{at+b}{ct+d}$ for some $a, b, c, d \in F$ with $ad - bc \neq 0$.

There are only countably many elements of $\mathbb{R}$ that are algebraic over $\mathbb{Q}$, since there are only countably many monic polynomials in $\mathbb{Q}[X]$, and each has only finitely many roots in $\mathbb{R}$. The next two problems exhibit uncountably many specific elements of $\mathbb{R}$ that are transcendental over $\mathbb{Q}$.

5.26. Let $\alpha \in \mathbb{R}$. Suppose that there is a sequence of rational numbers $r_1, r_2, \ldots$ with each r_j expressible as m_j/n_j with $m_j, n_j \in \mathbb{Z}$, $n_j \neq 0$, such that $0 < |\alpha - r_j| \leq 1/(j n_j^j)$ for each j. Prove that α is transcendental over $\mathbb{Q}$.

5.27.

(i) Let

$$\alpha = \sum_{n=1}^{\infty} 10^{-n!} = 0.11000100000000000000000010\ldots \in \mathbb{R}.$$

Use the preceding problem to prove that α is transcendental over $\mathbb{Q}$.

(ii) More generally, take any infinite sequence $\varepsilon_1, \varepsilon_2, \ldots$ such that $\varepsilon_i = 1$ when i is even and $\varepsilon_i = 0$ or 1 when i is odd. Prove that

$$\beta = \sum_{i=1}^{\infty} \varepsilon_i 10^{-n!} \in \mathbb{R}$$

is transcendental over $\mathbb{Q}$. Since there are uncountably many such infinite sequences of ε_i, this yields uncountably many elements of $\mathbb{R}$ transcendental over $\mathbb{Q}$.

5.28. Assume that $char(F) \neq 2$. Let field $L = F(t)$, where t is transcendental over F, and let $K = L\big(\sqrt{-(t^2+1)}\big)$ $(= F(t)(\gamma)$, where $\gamma^2 = -(t^2+1))$.

(i) Prove that F is algebraically closed in K, i.e., if $\alpha \in K$ is algebraic over F, then $\alpha \in F$.

(ii) Prove that $K = F(s)$ for some $s \in K$ iff there exist $\alpha, \beta \in F$ with $\alpha^2 + \beta^2 + 1 = 0$.

Algebraic independence. For fields $F \subseteq K$, elements $t_1, \ldots, t_n \in K$ are said to be *algebraically independent over* F if for any polynomial f in $F[X_1, \ldots, X_n]$, if $f(t_1, \ldots, t_n) = 0$ then $f = 0$ (or, equivalently, the evaluation map $\varepsilon_{F,t_1,\ldots,t_n} \colon F[X_1, \ldots, X_n] \to K$ is injective). When this occurs,

$$F[t_1, \ldots, t_n] \cong F[X_1, \ldots, X_n].$$

An infinite subset S of K is defined to be algebraically independent over F if each finite subset of S is algebraically independent over F.

5.29. Let $F \subseteq K$ be fields, and let $t_1, \ldots, t_n \in K$ for $n \in \mathbb{N}$. Prove that the following conditions are equivalent:

(a) $t_1, \ldots, t_n$ are algebraically independent over F.

(b) t_i is transcendental over $F(t_1, t_2, \ldots, t_{i-1})$ for $i = 1, 2, \ldots, n$.

(c) t_i is transcendental over $F(t_1, t_2, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$ for $i = 1, 2, \ldots, n$.

5.30. *Algebraic dependence.* Let $F \subseteq K$ be fields. Define a relation $\prec$, called *algebraic dependence over* F, between elements $s \in K$

and subsets T of K by

$$s \prec T \text{ iff } s \text{ is algebraic over } F(T). \tag{5.12}$$

Prove that $\prec$ is a dependence relation as defined on p. 130.

Transcendence bases and transcendence degree. Note that a subset I of K is independent for this relation $\prec$ of algebraic dependence iff I is algebraically independent over F, as defined above. Also subset U spans K for $\prec$ iff K is algebraic over $F(U)$. A base of B of K for algebraic dependence over F is called a *transcendence base of K over F*. Thus, B is a transcendence base iff B is algebraically independent over F and K is algebraic over $F(B)$. By problem 4.5, transcendence bases exist and every transcendence base of K over F has the same cardinality. The *transcendence degree of K over F*, denoted $trdeg_{K/F}$, is defined to be the cardinality of any transcendence base of K over F.

5.31. Let $F \subseteq L \subseteq K$ be fields. Prove that

$$trdeg_{K/F} = trdeg_{L/F} + trdeg_{K/L}.$$

5.32. Let $F \subseteq K$ be fields and let $\alpha, t \in K$ with α algebraic over F and t transcendental over F. Prove that $m_{F(t),\alpha} = m_{F,\alpha}$.

5.33. Let $F \subseteq L \subseteq K$ be fields, and suppose that K is finitely generated over F, i.e., $K = F(\alpha_1, \ldots, \alpha_n)$ for some $n \in \mathbb{N}$ and $\alpha_i \in K$. Prove that L is finitely generated over F. (Hint: Reduce to the case where L is algebraic over F.)

5.34. Let $R \subseteq T$ be integral domains. Suppose that R is a UFD with infinitely many nonassociate irreducibles, and that

$$T = R[t_1, \ldots, t_n] \quad \text{with each } t_i \text{ algebraic over } q(R).$$

Prove that T is not a field. (Hint: The idea here is that there are "too many inverses" of different elements of R, and not all of them are obtainable in $R[t_1, \ldots, t_n]$. Prove that there is $s \in R \setminus \{0\}$ such that each t_i is a root of some monic polynomial in $R[1/s][X]$, and that $T[1/s]^* \cap R[1/s] = R[1/s]^*$. Recall problems 3.75 and 3.28.)

5.35. *Zariski's Nullstellensatz.* Let $F \subseteq K$ be fields with $K = F[\alpha_1, \ldots, \alpha_n]$. Prove that K is algebraic over F. This is known as Zariski's form of the Nullstellensatz. (Hint: $\{\alpha_1, \ldots, \alpha_n\}$ contains a transcendence base of K over F. Apply the preceding problem.)

5.36. *Weak Nullstellensatz.* Let C be an algebraically closed field. Note that for any $a_1, \ldots, a_n \in C$, the ideal $(X - a_1, \ldots, X - a_n)$ of the polynomial ring $C[X_1, \ldots, X_n]$ is a maximal ideal, since it is the kernel of the surjective evaluation homomorphism

$$\varepsilon_{C,a_1,\ldots,a_n} : C[X_1, \ldots, X_n] \longrightarrow C$$

(recall problem 3.50). Now prove conversely that for every maximal ideal M of $C[X_1, \ldots, X_n]$,

$$M = (X - a_1, \ldots, X - a_n) \quad \text{for some } a_1, \ldots, a_n \in C.$$

This is known as the weak Nullstellensatz. (Hint: Apply the preceding problem to $C[X_1, \ldots, X_n]/M = C[t_1, \ldots, t_n]$, where $t_i = X_i + M$.)

5.37. *Hilbert's Nullstellensatz.* Let C be an algebraically closed field, and take any polynomials $f_1, f_2, \ldots, f_m, g$ in $C[X_1, \ldots, X_n]$. Prove that the following conditions are equivalent:

(a) For any $a_1, \ldots, a_n \in C$, if $f_i(a_1, \ldots, a_n) = 0$ for each i, then $g(a_1, \ldots, a_n) = 0$.

(b) g^r lies in the ideal $(f_1, \ldots, f_m)$ of $C[X_1, \ldots, X_n]$, for some $r \in \mathbb{N}$.

(c) The ring $C[X_1, \ldots, X_n, Y]/(f_1, \ldots, f_m, (1 - gY))$, is trivial.

(Hint: The ring of (c) is nontrivial iff it has a maximal ideal, by problem 3.43. Use problems 3.48(iv) and 5.36.) The equivalence of (a) and (b) is *Hilbert's Nullstellensatz*, which is a cornerstone of algebraic geometry.

5.4. Criteria for irreducibility of polynomials

5.38. Let R be a UFD with quotient field K. Let f be a nonconstant polynomial in $R[X]$ (i.e., $deg(f) \geq 1$). Prove that if f is reducible in $K[X]$, then there are $g, h \in R[X]$ with $f = gh$ and $deg(g) < deg(f)$ and $deg(h) < deg(f)$. (See problems 3.79 and 3.80(i) above.)

5.39. Let

$$f = X^4 + 18X^3 + 11X^2 + 10X + 9 \in \mathbb{Z}[X].$$

Prove that f is irreducible in $\mathbb{Q}[X]$. (Hint: Consider the images of f in $\mathbb{Z}_2[X]$ and in $\mathbb{Z}_3[X]$.)

Recall *Eisenstein's Irreducibility Criterion*: Let R be a UFD with quotient field K, and let π be irreducible in R. In $R[X]$ let $f = \sum_{i=0}^{n} c_i X^i$, with $n \geq 1$. Suppose that π divides each of $c_0, c_1, \ldots, c_{n-1}$ in R, but $\pi \nmid c_n$ and $\pi^2 \nmid c_0$. Then f is irreducible in $K[X]$.

(Proof sketch: If f is reducible in $K[X]$, then $f = gh$ in $R[X]$ with $deg(g) < deg(f)$ and $deg(h) < deg(f)$ (see problem 5.38). Let $\overline{f}$, $\overline{g}$, $\overline{h}$, be the images of f, g, h in $(R/(\pi))[X]$ Since $\overline{f} = \overline{c_n} X^n = \overline{g}\,\overline{h}$ and $R/(\pi)$ is an integral domain, it follows that $\overline{g}(0) = \overline{h}(0) = 0$ contradicting $\pi^2 \nmid c_0$.)

5.40. Let R be a UFD with quotient field K, and let π be irreducible in R. Let

$$f = (X - c)^n + \pi g \in R[X],$$

where $c \in R$ and $g \in R[X]$ with $\pi \nmid g(c)$ in R and $deg(g) \leq n$. Prove that f is irreducible in $K[X]$.

5.41. Prove that $X^{27} - 4$ is irreducible in $\mathbb{Q}[X]$.

5.42. Let f be nonconstant in $F[X]$, and suppose that for every field $K \supseteq F$ if f has a root in K, then f splits over K. Prove that every irreducible factor of f in $F[X]$ has the same degree.

5.43. Let $f \in F[X]$ with $deg(f)$ a prime number. Suppose that f splits over every field $K \supseteq F$ containing a root of f. Prove that either f is irreducible in $F[X]$ or f splits over F.

Example 5.44. Here are some examples where the preceding problem applies. Let p be a prime number.

(i) Suppose that $char(F) \neq p$ and F contains a primitive p-th root of unity ω, i.e., $\omega^p = 1$ but $\omega \neq 1$. If $a \in F \setminus F^p$, then

$$f = X^p - a$$

is irreducible in $F[X]$. (If α is a root of f in a field $K \supseteq F$, then $\alpha, \omega\alpha, \ldots, \omega^i\alpha, \ldots, \omega^{p-1}\alpha$ are p distinct roots of f in K; so f splits over K.)

(ii) Suppose that $char(F) = p$. If $a \in F \setminus F^p$, then

$$f \;=\; X^p - a$$

is irreducible in $F[X]$. (If α is a root of f in a field $K \supseteq F$, then in $K[X]$, $f = X^p - \alpha^p = (X - \alpha)^p$.)

(iii) Suppose that $char(F) = p$ and let

$$f \;=\; X^p - X - a \;\in\; F[X].$$

Then either f is irreducible in $F[X]$ or f splits over F. (If α is a root of f in a field $K \supseteq F$, then $\alpha, \alpha + 1, \ldots, \alpha + i, \ldots, \alpha + p - 1$ are p distinct roots of f in K; hence, f splits over K.)

5.45. Let p be a prime number, and let F be a field with $char(F) \neq p$. Let

$$f \;=\; X^p - a \;\in\; F[X].$$

Prove that either f is irreducible or f has a root in F. (But f need not split over F.)

5.46. Let

$$f \;=\; X^7 + 14x^4 + 6X + 9 \;\in\; \mathbb{Z}[X].$$

Prove that f is irreducible in $\mathbb{Q}[X]$. (Hint: Consider the image of f in $\mathbb{Z}_7[X]$ and use Example 5.44(iii).)

5.47. Suppose that $char(F) = p \neq 0$, and let

$$f \;=\; X^{p^n} - a \in F[X], \quad \text{for some } n \in \mathbb{N}.$$

Observe that if β is a root of f in some field K containing F, then $\beta^{p^n} = a$, so $f = X^{p^n} - \beta^{p^n} = (x - \beta)^{p^n}$ in $K[X]$, showing that β is the only root of f in K.

(i) Prove that the irreducible factorization of f in $F[X]$ has the form $f = (X^{p^k} - c)^{p^{n-k}}$ for some $c \in F$.

(ii) Prove that f is irreducible in $F[X]$ iff $a \notin F^p$.

5.48. Let $m, n \in \mathbb{N}$ with $gcd(m, n) = 1$, and let $a \in F$. Prove that if $X^m - a$ and $X^n - a$ are irreducible in $F[X]$, then $X^{mn} - a$ is irreducible in $F[X]$.

5.49. This problem gives Kronecker's algorithm for factoring a polynomial in $\mathbb{Q}[X]$ (so also testing for irreducibility). Let $f \in \mathbb{Q}[X]$ with $deg(f) \geq 2$. Since up to a constant factor f is primitive in $\mathbb{Z}[X]$, we may assume at the outset that $f \in \mathbb{Z}[X]$ and f is primitive. Let m be the integer with $m \leq deg(f)/2 < m + 1$. Note that if f is reducible in $\mathbb{Q}[X]$, then there is $g \in \mathbb{Z}[X]$ such that $g|f$ in $\mathbb{Z}[X]$ and $deg(g) \leq m$ (recall problem 5.38 above). Take any distinct $n_1, n_2, \ldots, n_{m+1} \in \mathbb{Z}$, and let $f(n_i) = s_i \in \mathbb{Z}$. If some $s_i = 0$, then $(X - n_i)|f$ in $\mathbb{Z}[X]$, so f is reducible. Thus, we may assume that each $s_i \neq 0$. Now, for each i, choose $d_i \in \mathbb{Z}$ with $d_i|s_i$. By Lagrange interpolation (see problem 4.11), there is a unique $h \in \mathbb{Q}[X]$ with $deg(h) \leq m$ and $h(n_i) = d_i$ for all i. One can check whether $h|f$ by polynomial long division. There are only finitely many choices for the d_i, so only finitely many of these polynomials h. Prove that if there is $g \in \mathbb{Z}[X]$ such that $deg(g) \leq m$ and $g|f$, then g must be one of the h's for some choice of the d_i. This gives an algorithm for finding factors if f in just finitely many steps, since there are only finitely many choices for h, and each can be checked in turn. However, this is computationally extremely inefficient, both because it involves testing so many cases, and also because it requires the computationally difficult (but finite) task of determining all the integral divisors of each s_i.

5.5. Splitting fields, normal field extensions, and Galois groups

Splitting fields. Take any nonconstant $f \in F[X]$. A field E containing F is called a *splitting field of f over F* if both

(i) f splits over E; and

(ii) $E = F(\alpha_1, \alpha_2, \ldots, \alpha_k)$, where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are all the roots of f in E.

Equivalently, field $E \supseteq F$ is a splitting field of f over F if f splits over E but f does not split over any proper subfield of E. (Thus, the expression "minimal splitting field" would be more descriptive, but "splitting field" is the standard terminology.) The existence of splitting fields follows by Kronecker's Theorem. See Note 5.2 above.

5.50. Let $f \in F[X]$ and let $n = deg(f) \in \mathbb{N}$, and let E be a splitting field of f over F. Clearly (from the proof in Note 5.2), $[E{:}F] \leq n!$. Now prove that $[E{:}F] \big| n!$.

5.51. Suppose that $char(F) \neq 2$, and let

$$f \;=\; X^4 + 2bX^2 + c \in F[X].$$

Prove that f is reducible in $F[X]$ iff either of the following two conditions holds:

(i) $b^2 - c \in F^2$;

(ii) $c \in F^2$, say $c = \gamma^2$, and one of $2(\gamma - b)$ or $-2(\gamma + b)$ lies in F^2.

(This amounts to: The only possible factorizations of f into factors of degree 2 are the ones given by completing the the square: $f = (X^2 + b)^2 - (b^2 - c)$, which factors if $b^2 - c \in F^2$; but also, if $c = \gamma^2 \in F$, then $f = (X^2 \pm \gamma)^2 - 2(\pm\gamma - b)X^2$, which factors if it is a difference of squares.)

5.52. As in the preceding problem, suppose that $char(F) \neq 2$ and let

$$f \;=\; X^4 + 2bX^2 + c \;\in\; F[X].$$

Assume that f is irreducible. Let E be a splitting field of f over F. Note that f has four distinct roots in E (by direct calculation or by the Derivative Test; see p. 217 below), and that if α is a root of f then $-\alpha$ is also a root. Thus, the roots of f can be described as $\pm\alpha$ and $\pm\beta$, where $\beta \neq \pm\alpha$.

(i) Prove that $[E{:}F] = 4$ or $= 8$.

(ii) Prove that $[E{:}F] = 8$ iff $c \notin F^2$ and $c(b^2 - c) \notin F^2$. (Hint: Let $L = F(\alpha^2)$. Verify that $L = F(\beta^2) = F\big(\sqrt{b^2 - c}\,\big)$. To determine whether $[E{:}L] = 4$ or $= 2$, apply problem 5.5(ii).)

F-homomorphisms. If F and F' are fields, a homomorphism $\tau\colon F \to F'$ is understood to be a ring homomorphism. Then τ is necessarily injective, as F is a field. Likewise, isomorphisms and automorphisms of fields are understood to be ring isomorphisms and

automorphisms. Any homomorphism $\tau\colon F \to F'$ induces a ring homomorphism, also denoted τ, mapping $F[X]$ to $F'[X]$, given by

$$\tau(\textstyle\sum c_i X^i) \;=\; \sum \tau(c_i) X^i.$$

Note that

$$\text{if } \alpha \in F \text{ is a root of } f \in F[X], \text{ then } \tau(\alpha) \text{ is a root of } \tau(f), \quad (5.13)$$

since $0_{F'} = \tau(0_F) = \tau(f(\alpha)) = \tau(f)(\tau(\alpha))$.

If L and K are extension fields of F, then a homomorphism $\sigma\colon L \to K$ is called an *F-homomorphism* if $\sigma(c) = c$ for all $c \in F$. F-isomorphisms and F-automorphisms are defined analogously.

Recall the *Isomorphism Extension Theorem* (IET): Let F_1 and F_2 be fields, and let $\rho\colon F_1 \to F_2$ be an isomorphism. Let $f_1 \in F_1[X]$ with $deg(f_1) \geq 1$, and let $f_2 = \rho(f_1) \in F_2[X]$. Let E_i be a splitting field of f_i over F_i for $i = 1, 2$. Then there is an isomorphism $\tau\colon E_1 \to E_2$ with $\tau|_{F_1} = \rho$.

(Proof sketch: If $E_1 = F_1$, then f_2 splits over F_2, so $E_2 = F_2$ and we take $\tau = \rho$. If $E_1 \neq F_1$, let g_1 be an irreducible factor of f_1 in $F_1[X]$ with $deg(g_1) \geq 2$, and let $g_2 = \rho(g_1)$, which is an irreducible factor of f_2 in $F_2[X]$. Let α_i be a root of g_i in E_i for $i = 1, 2$. Let $\sigma\colon F_1(\alpha_1) \to F_2$ be the composition of isomorphisms

$$F_1(\alpha_1) \xrightarrow{\cong} F_1[X]/(g_1) \xrightarrow{\cong} F_2[X]/(g_2) \xrightarrow{\cong} F_2(\alpha_2).$$

The first and last isomorphisms here are given by (5.6) and the middle one is induced by ρ. Then, $\sigma|_{F_1} = \rho$. Since E_i is a splitting field of f_i over $F_i(\alpha_i)$, by induction on $[E_1\colon F_1]$ there is an isomorphism $\tau\colon E_1 \to E_2$ with $\tau|_{F_1(\alpha_1)} = \sigma$; so, $\tau|_{F_1} = \rho$.) See, e.g., Dummit & Foote [**5**, Th. 27, p. 541] or Cox [**4**, Th. 5.1.6, pp. 103–104] for more detailed proofs of this crucial theorem.

Note 5.53. *Uniqueness of splitting fields.* One immediate but significant consequence of the Isomorphism Extension Theorem is the uniqueness of splitting fields up to isomorphism: If E_1 and E_2 are each splitting fields of $f \in F[X]$ over F, then E_1 and E_2 are F-isomorphic. Proof: Apply the IET with $F_1 = F_2 = F$, $\rho = id_F$, and $f_1 = f_2 = f$.

5.54. Let f be irreducible in $F[X]$, let E be a splitting field of f over F, and let α, β be roots of f in E. We know that E is unique up to F-isomorphism, and that by (5.6), $F(\alpha)$ is F-isomorphic to $F(\beta)$. Now give an example of F and irreducible f with four of its distinct roots $\alpha_1, \alpha_2, \beta_1, \beta_2$ in E such that $F(\alpha_1, \alpha_2) \ncong F(\beta_1, \beta_2)$.

Splitting fields of families of polynomials. Let $\{f_i\}_{i \in I}$ be a family of nonconstant polynomials in $F[X]$. A field $K \supseteq F$ is called a *splitting field of the f_i over F* if each f_i splits over K and K is generated over F by all the roots in K of all the f_i. Thus, there is no proper subfield of K containing F over which all the f_i split. Note that such a splitting field always exists: Let A be an algebraic closure of F, and let K be the subfield of A generated by F and all the roots in A of all the f_i.

5.55. Prove the following generalization of the Isomorphism Extension Theorem: Let F_1 and F_2 be fields, and let $\rho\colon F_1 \to F_2$ be an isomorphism. Let $\{f_i\}_{i \in I}$ be a family of nonconstant polynomials in $F_1[X]$ with and let $g_i = \rho(f_i) \in F_2[X]$ for all $i \in I$. Let E_1 be a splitting field of $\{f_i\}_{i \in I}$ over F_1, and E_2 a splitting field of $\{g_i\}_{i \in I}$ over F_2. Prove that there is an isomorphism $\tau\colon E_1 \to E_2$ with $\tau|_{F_1} = \rho$. (Hint: Apply Zorn's Lemma to the set $\mathcal{S}$ of pairs (L, σ), where L is a field with $F_1 \subseteq L \subseteq E_1$ and $\sigma\colon L \to E_2$ is a ring homomorphism with $\sigma|_{F_1} = \rho$. The set $\mathcal{S}$ is partially ordered by $(L, \sigma) \leq (L', \sigma')$ iff L is a subfield of L' and $\sigma'|_L = \sigma$.)

Note 5.56. *Uniqueness of splitting fields and algebraic closures.* It follows from the generalized Isomorphism Extension Theorem that splitting fields are unique up to isomorphism: If $\{f_i\}_{i \in I}$ is any family of nonconstant polynomials in $F[X]$ and E_1 and E_2 are each splitting fields of the f_i over F, then E_1 and E_2 are F-isomorphic. This is proved just as in Note 5.53 above. It follows that if fields A_1 and A_2 are two algebraic closures of F, then A_1 and A_2 are F-isomorphic. For, each A_i is a splitting field over F of the set of all nonconstant polynomials in $F[X]$.

Galois group. Let $F \subseteq K$ be fields. The *Galois group of K over F* is

$$\mathcal{G}(K/F) = \{\tau\colon K \to K \mid \tau \text{ is an isomorphism and } \tau|_F = id_F\}, \tag{5.14}$$

the set of F-automorphisms of K, which is a group with the operation of composition of functions. Its identity element is the identity map id_K.

Note that if $[K{:}F] < \infty$, then $\big|\mathcal{G}(K/F)\big| < \infty$, since if $K = F(\alpha_1, \ldots, \alpha_k)$, then $\tau \in \mathcal{G}(K/F)$ is determined by $\tau(\alpha_1), \ldots, \tau(\alpha_k)$, and each $\tau(\alpha_i)$ is a root of m_{F,α_i} (see (5.13)). A better upper bound on the order of the Galois group is given by

$$\big|\mathcal{G}(K/F)\big| \leq [K{:}F]. \tag{5.15}$$

Proof: Let $G = \mathcal{G}(K/F)$. Argue by induction on $[K{:}F]$. If $[K{:}F] = 1$, then $K = F$ and $\mathcal{G}(K/F) = \{id_F\}$. If $[K{:}F] > 1$, take any $\alpha \in K \setminus F$. Let $L = F(\alpha)$, and let

$$H = \mathcal{G}(K/L) = \{\sigma \in G \mid \sigma(\alpha) = \alpha\},$$

which is a subgroup of G. By induction, $|H| \leq [K{:}L]$. For each $\sigma \in G$, $\sigma(\alpha)$ is a root of $m_{F,\alpha}$ (see (5.13)). Let

$$\{\sigma(\alpha) \mid \sigma \in G\} = \{\alpha_1, \ldots, \alpha_k\} \quad \text{with the } \alpha_i \text{ distinct.}$$

Then,

$$k \leq \text{the number of roots of } m_{F,\alpha} \text{ in } K \leq deg(m_{F,\alpha}) = [L{:}F].$$

For any $\sigma \in G$, there is a j with $\sigma(\alpha) = \tau_j(\alpha)$. Then, $\tau_j^{-1}\sigma \in H$, as $\tau_j^{-1}\sigma(\alpha) = \alpha$. Hence, σ lies in the coset $\tau_j H$ of H in G. This shows that $G = \tau_1 H \cup \ldots \cup \tau_k H$, and

$$|G| \leq k\,|H| \leq [L{:}F] \cdot [K{:}L] = [K{:}F].$$

5.57. Let f be irreducible in $F[X]$, and let E be a splitting field of f over F. Let α and β be roots of f in E. Prove that there is $\tau \in \mathcal{G}(K/F)$ with $\tau(\alpha) = \beta$. (Hint: Prove that there is an F-isomorphism $\rho\colon F(\alpha) \to F(\beta)$ with $\rho(\alpha) = \beta$. See (5.6). Apply the IET.)

Note 5.58. The *Four Field Theorem* says: Let $F \subseteq L \subseteq E \subseteq K$ be fields with E a splitting field over F of some $f \in F[X]$. Let $\rho\colon L \to K$ be an F-homomorphism. Then, $\rho(L) \subseteq E$, and there is $\tau \in \mathcal{G}(E/F)$ with $\tau|_L = \rho$.

Proof: We have $E = F(\alpha_1, \ldots, \alpha_k)$, where $\alpha_1, \ldots, \alpha_k$ are all the distinct roots of f in E. Let

$$M_1 = L(\alpha_1, \ldots, \alpha_k) = E \quad \text{and} \quad M_2 = \rho(L)(\alpha_1, \ldots, \alpha_k) \subseteq K.$$

Let $f_1 = f$ and $f_2 = \rho(f_1) = f$. Then, M_1 is a splitting field of f_1 over L and M_2 is a splitting field of f_2 over $\rho(L)$. By the IET (applied with ground fields L and $\rho(L)$), there is an isomorphism $\tau\colon M_1 \to M_2$ with $\tau|_L = \rho$. Hence, $\tau|_F = \rho|_F = id_F$. For any i, we have $\tau(\alpha_i)$ is a root of $\tau(f) = f$ (see (5.13)), so $\tau(\alpha_i)$ must be one of the α_j. Hence, τ maps $\{\alpha_1, \ldots, \alpha_k\}$ into itself; as τ is injective, it must map this set onto itself. Hence,

$$\begin{aligned}\tau(E) &= \tau\big(F(\alpha_1, \ldots, \alpha_k)\big) = \tau(F)\big(\tau(\alpha_1), \ldots, \tau(\alpha_k)\big) \\ &= F(\alpha_1, \ldots, \alpha_k) = E.\end{aligned}$$

Thus, $\tau \in \mathcal{G}(E/F)$, and, as $L \subseteq E$, we have $\tau(L) \subseteq \tau(E) = E$.

5.59. Let E be a splitting field over F of some nonconstant $f \in F[X]$, and take any irreducible $g \in F[X]$. Prove that if g has a root in E then g splits over E. (Hint: Let K be a splitting field of g over E, and let α, β be roots of g in K with $\alpha \in E$. Apply the Four Field Theorem with $L = F(\alpha) \cong F(\beta)$.)

Normal field extensions. Let $F \subseteq E$ be fields with $[E{:}F] < \infty$. Then E is said to be *normal over* F if every irreducible polynomial in $F[X]$ with a root in E splits over E. (Equivalently, E is normal over F iff $m_{F,\alpha}$ splits over E for each $\alpha \in E$.) The preceding problem shows the remarkable fact that E is normal over F iff E is a splitting field over F of some nonconstant polynomial in $F[X]$. (For "only if," note that if E is normal over F and $E = F(\alpha_1, \ldots, \alpha_k)$, then E is a splitting field over F of the product $m_{F,\alpha_1} m_{F,\alpha_2} \ldots m_{F,\alpha_n}$.)

5.60. *Normal closure.* Let $F \subseteq L$ be fields with $[L{:}F] < \infty$, say $L = F(\alpha_1, \ldots, \alpha_n)$. Let $f = m_{F,\alpha_1} m_{F,\alpha_2} \ldots m_{F,\alpha_n} \in F[X]$, and

let E be a splitting field of f over L. The field E is called a *normal closure of* L *over* F; this problem justifies the terminology.

(i) Prove that E is normal over F.

(ii) Let M be any extension field of E, and let K be a field with $L \subseteq K \subseteq M$. Prove that if K is normal over F then $E \subseteq K$. Thus, E is the minimal extension of L in K that is normal over F.

5.61. Let $F \subseteq L \subseteq K$ be fields with L normal over F and K normal over L. Then, K need not be normal over F. (See problem 5.63(ii) below for an example of this.) But we can build from K a normal extension of F containing K as follows: Let K be a splitting field over L of $f \in L[X]$. Let $g = \prod_{\tau \in \mathcal{G}(L/F)} \tau(f) \in L[X]$, and let E be a splitting field of g over K. Prove that E is normal over F.

5.62. Let $F \subseteq E$ be fields with E normal over F, and take any irreducible $f \in F[X]$. Let g and h be monic irreducible factors of f in $E[X]$. Prove that there is $\tau \in \mathcal{G}(E/F)$ with $\tau(g) = h$. (It follows that all the irreducible factors of f in E have the same degree.)

5.63. Suppose that $char(F) \neq 2$, and let

$$f \;=\; X^4 + 2bX^2 + c$$

be irreducible in $F[X]$. (The conditions for this are given in problem 5.51.) Let E be a splitting field of f over F. We know from problem 5.52 that $[E{:}F] = 4$ or $= 8$, and that we can describe the four different roots of f in E as $\pm\alpha$ and $\pm\beta$. Note that the IET and problem 5.57 can be used to build elements of $\mathcal{G}(E/F)$.

(i) Prove that if $[E{:}F] = 8$, then

$$\mathcal{G}(E/F) \;\cong\; D_4,$$

the dihedral group of order 8.

(ii) Suppose $[E{:}F] = 8$. Prove that $F(\alpha)$ is normal over $F(\alpha^2)$ and $F(\alpha^2)$ is normal over F, but $F(\alpha)$ is not normal over F.

(iii) Suppose for the rest of this problem that $[E{:}F] = 4$. By problem 5.52(ii), $c \in F^2$ or $c(b^2 - c) \in F^2$, but not both, since $b^2 - c \notin F^2$ by problem 5.51, as f is irreducible. Prove that $|\,\mathcal{G}(E/F)| = 4$.

(iv) Prove that if $c \in F^2$, then

$$\mathcal{G}(E/F) \cong C_2 \times C_2$$

(the noncyclic group of order 4). (Hint: Observe that $\alpha^2\beta^2 = c \in F^2$, hence $\alpha\beta \in F$.) Also, find $d, e \in F$ such that $E = F\big(\sqrt{d}\,,\sqrt{e}\,\big)$.

(v) Prove that if $c(b^2 - c) \in F^2$, then

$$\mathcal{G}(E/F) \cong C_4$$

(the cyclic group of order 4). (Hint: Show that $\alpha\beta(\alpha^2 - \beta^2) \in F$.)

5.64. Let $K = F(t)$, where t is transcendental over F.

(i) Take any $a, b, c, d \in F$ with $ad - bc \neq 0$, and let $s = \frac{at+b}{ct+d} \in K$. Prove that there is $\tau \in \mathcal{G}(K/F)$ with $\tau(t) = s$, and that there is only one such τ. (Recall from problem 5.25 that s is transcendental over F and that $F(s) = K$.)

(ii) For $A = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right) \in GL_2(F)$, let τ_A be the element of $\mathcal{G}(K/F)$ with $\tau_A(t) = \frac{at+b}{ct+d}$, as in part (i). Prove that the map $\psi\colon GL_2(F) \to \mathcal{G}(K/F)$ given by $A \mapsto \tau_A$ is a surjective group homomorphism.

Note that

$$ker(\psi) = \left\{ \left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right) \mid a \in F^* \right\} = Z(GL_2(F)).$$

The factor group $GL_2(F)/Z(GL_2(F))$ is called the *projective linear group of degree* 2 *of* F, denoted $PGL_2(F)$. Thus, we have an isomorphism for t transcendental over F,

$$\mathcal{G}(F(t)/F) \cong PGL_2(F). \tag{5.16}$$

5.65. Prove that $\mathcal{G}(\mathbb{R}/\mathbb{Q}) = \{id_{\mathbb{R}}\}$.

5.66. Suppose that $char(F) = 0$, and let field $K = F(t)$, where t is transcendental over F. Let

$$L_1 = F(t^2) \qquad \text{and} \qquad L_2 = F(t^2 + t),$$

which are subfields of K. We know from problem 5.25 that $[K\colon L_i] = 2$ for $i = 1, 2$.

(i) Determine each automorphism in $\mathcal{G}(K/L_i)$ for $i = 1$ and $i = 2$. (Hint: Since $K = L_i(t)$, each $\sigma \in \mathcal{G}(K/L_i)$ is determined by its effect on t. To find possible $\sigma(t)$ first compute $m_{L_i,t}$.)

(ii) Use part (i) to show that $\left|\mathcal{G}(K/(L_1 \cap L_2))\right| = \infty$. Then deduce that $L_1 \cap L_2 = F$. (Recall (5.15). The last equality is difficult to prove without using the Galois group information.)

5.67. Let F, K, L_1, L_2 be as in the preceding problem, except assume now that $char(F) = p > 2$.

(i) Prove that $[K : L_1 \cap L_2] = 2p$, and find $s \in L_1 \cap L_2$ such that $L_1 \cap L_2 = F(s)$.

(ii) Prove that $\mathcal{G}(K/(L_1 \cap L_2)) \cong D_p$, the p-th dihedral group.

5.68. Let field $K = F(t)$, where t is transcendental over F. Prove *Lüroth's Theorem*: for any field L with $F \subseteq L \subseteq K$, there is $s \in K$ with $L = F(s)$. (Hint: If $L \neq F$, then t is algebraic over L. Let s be any coefficient of $m_{L,t}$ not lying in F. Prove that $L = F(s)$ by comparing $m_{L,t}$ and $m_{F(s),t}$.)

5.6. Separability and repeated roots

Formal derivative. For $f = c_n X^n + \ldots + c_i X^i + \ldots + c_0 \in F[X]$, the *formal derivative of* f is defined to be

$$f' = nc_n X^{n-1} + \ldots + ic_i X^{i-1} + \ldots + 2c_2 X + c_1 \in F[X]. \quad (5.17)$$

This is meaningful for any field F. Note that the formal derivative satisfies the following properties familiar for the usual derivative: for any $f, g \in F[X]$,

(i) *F-linearity*: for any $c, d \in F$, $(cf + dg)' = cf' + dg'$;

(ii) *product rule*: $(fg)' = f'g + fg'$;

(iii) *chain rule*: let $h = f(g(X))$; then $h' = f'(g(X)) \cdot g'$.

Note also that if $char(F) = 0$, then $f' = 0$ iff f is a constant polynomial. But if $char(F) = p \neq 0$, then $f' = 0$ iff $f \in F[X^p]$.

Repeated roots. Let f be nonconstant in $F[X]$, and let α be a root of f in some extension field of F. The *multiplicity* of α as a root of f is the $m \in \mathbb{N}$ such that $(X-\alpha)^m | f$ but $(X-\alpha)^{m+1} \nmid f$ in $F(\alpha)[X]$. Then α is said to be a *simple root* of f if $m = 1$. But α is a *repeated root* of f if $m > 1$.

Recall the *Derivative Test* for repeated roots: For any nonconstant $f \in F[X]$, f has no repeated roots in any field containing F iff $gcd(f, f') \sim 1$ in $F[X]$. See any text covering field theory for a proof of the Derivative Test. (Proof idea: α is a repeated root of f iff $f(\alpha) = 0$ and $f'(\alpha) = 0$. This occurs iff $m_{F,\alpha} | f$ and $m_{F,\alpha} | f'$ in $F[X]$.)

Separability. A nonconstant polynomial $f \in F[X]$ is said to be *separable* if f has no repeated roots in any field containing F. By the Derivative Test, f is separable iff $gcd(f, f') \sim 1$ in $F[X]$. Note that if f is irreducible in $F[X]$, then either $gcd(f, f') \sim 1$ or $gcd(f, f') \sim f$. The latter case occurs iff $f | f'$, iff $f' = 0$ (as $deg(f') < deg(f)$ if $f' \neq 0$), iff $char(F) = p \neq 0$ and $f \in F[X^p]$. Thus, when f is irreducible, f is separable iff either (i) $char(F) = 0$; or (ii) $char(F) = p \neq 0$ and $f \notin F[X^p]$.

An element α in a field $K \supseteq F$ is said to be *separable over* F just when α is algebraic over F and the polynomial $m_{F,\alpha}$ is separable. In view of the comments just above, if $char(F) = 0$, then α is separable over F whenever α is algebraic over F; the concept of separability is then not needed. A field extension L of F is said to be *separable over* F if every element of L is separable over F. If $char(F) = 0$, then L is separable over F iff L is algebraic over F.

5.69. Suppose that $char(F) = p \neq 0$, and take α in some field K containing F, with α algebraic but not separable over F. Let $f = m_{F,\alpha}$. We know that $f' = 0$, so $f = g(X^p)$ for some $g \in F[X]$.

(i) Prove that g is irreducible in $F[X]$.

(ii) Prove that $f = h(X^{p^n})$ for some irreducible separable h in $F[X]$ and some $n \in \mathbb{N}$. (Hence, α^{p^n} is separable over F, even though α is not.)

(iii) Prove that every root of $m_{F,\alpha}$ in a field over which it splits occurs with multiplicity p^n (for the n of part (ii)).

Pure inseparability. Let $F \subseteq K$ be fields and suppose that $char(F) = p \neq 0$. An element $\alpha \in K$ is said to be *purely inseparable over* F if $\alpha^{p^n} \in F$ for some $n \in \mathbb{N}$. Equivalently, for α algebraic over F, α is purely inseparable over F iff α is the only root of $m_{F,\alpha}$ in any field containing α, iff $m_{F,\alpha} = X^{p^k} - c$ for some $c \in F$ (see problem 5.47). The field K is said to be *purely inseparable over* F if each element of K is purely inseparable over F. Note that if $[K{:}F] < \infty$ and K is purely inseparable over F, then K is normal over F and $[K{:}F]$ is a power of p.

5.70. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. This problem and the next two give an approach to separability of K over F that is independent of the main results of Galois theory. For this, let $E \supseteq K$ be a field with E normal over F, e.g., a normal closure of K over F (see problem 5.60). Let L be a field with $F \subseteq L \subseteq K$, and let $\iota\colon L \to E$ be an F-homomorphism. By the Four Field Theorem 5.58 (for the fields $F \subseteq L \subseteq E \subseteq E$) there is $\tau \in \mathcal{G}(E/F)$ with $\tau|_L = \iota$. We count the number of extensions of ι to K: Let

$$\#(K/L,\iota) = \big|\{\theta \mid \theta\colon K \to E \text{ is an } F\text{-homomorphism and } \theta|_L = \iota\}\big|.$$

Each such θ extends to some $\sigma \in \mathcal{G}(E/F)$. Thus,

$$1 \leq \#(K/L,\iota) \leq \big|\,\mathcal{G}(E/F)\big| < \infty.$$

(i) Take any $\alpha \in K$. Note that as E is normal over F, it is also normal over L and over $\iota(L)$. Hence, $m_{L,\alpha}$ splits over E. Since $\iota(m_{L,\alpha}) = m_{\iota(L),\tau(\alpha)}$ for any $\tau \in \mathcal{G}(E/F)$ with $\tau|_L = \iota$, $\iota(m_{L,\alpha})$ also splits over E. Prove that

$$\begin{aligned}\#(L(\alpha)/L,\iota) &= \text{the number of distinct roots of } \iota(m_{L,\alpha}) \text{ in } E\\ &= \text{the number of distinct roots of } m_{L,\alpha} \text{ in } E\\ &\leq [L(\alpha){:}L].\end{aligned}$$

(ii) Let M be a field with $L \subseteq M \subseteq K$, and let $\theta_1, \ldots, \theta_\ell$ be the distinct extensions of ι to F-homomorphisms $M \to E$; so, $\ell = \#(M/L,\iota)$. Prove that

$$\#(K/L,\iota) = \sum_{j=1}^{\ell} \#(K/M,\theta_j).$$

(iii) Deduce (by induction on $[K{:}L]$) that $\#(K/L,\iota) \leq [K{:}L]$.

(iv) Prove that the following conditions are equivalent:
   (a) $\#(K/F, id_F) = [K{:}F]$.
   (b) K is separable over F.
   (c) For some $\alpha_1 \ldots, \alpha_n \in K$ with $K = F(\alpha_1, \ldots, \alpha_n)$, each α_i is separable over $F(\alpha_1, \ldots, \alpha_{i-1})$.

5.71. *Separable closure.* Let $F \subseteq K$ be fields with K algebraic over F.

(i) Let $\alpha, \beta \in K$. Prove that if α and β are separable over F then the field $F(\alpha, \beta)$ is separable over F. (Use the preceding problem.)

(ii) Let
$$S = \{\alpha \in K \mid \alpha \text{ is separable over } F\}.$$
This S is called the *separable closure of F in K*. Prove that S is a subfield of K with $F \subseteq S$.

(iii) Prove that if $\alpha \in K$ and α is separable over S, then $\alpha \in S$.

(iv) Suppose that $char(F) = p \neq 0$. Prove that if $\gamma \in K$ then $\gamma^{p^k} \in S$ for some $k \in \mathbb{N}$. Hence, K is purely inseparable over S.

(v) Suppose that $[K{:}F] < \infty$. Prove that, in the notation of the preceding problem, $\#(K/F, id_F) = [S{:}F]$.

5.72. Let $F \subseteq L \subseteq K$ be fields. Prove that if L is separable over F and K is separable over L, then K is separable over F.

5.73. Let $F \subseteq L$ be fields with $[L{:}F] < \infty$ and L separable over F. Let E be a normal closure of L over F. Prove that E is separable over F.

5.74. *Purely inseparable closure.* Let $F \subseteq K$ be fields of nonzero characteristic p. Let
$$I = \{\alpha \in K \mid \alpha \text{ is purely inseparable over } F\}.$$
This I is called the *purely inseparable closure of F in K*.

(i) Prove that I is a subfield of K, with $F \subseteq I$.

(ii) Prove that if $\beta \in K$ is purely inseparable over I, then $\beta \in I$.

5.75. Take any $n \in \mathbb{N}$, and let F be a field with $char(F) = 0$ or $char(F) = p$ with $p \nmid n$. Let $f = X^n - 1 \in F[X]$, let K be a splitting field of f over F. Let U be the set of roots of f in K, i.e.,

$$U = \{\zeta \in K \mid \zeta^n = 1\}.$$

The Derivative Test applied to f shows that $|U| = n$. Since U is clearly a finite subgroup of K^*, it is a cyclic group (see problem 3.32). The elements of U are called the *n-th roots of unity* in K. An element of U of order n (i.e., a generator of the cyclic group U) is called a *primitive n-th root of unity* in K. Let ω be any primitive n-th root of unity in K.

(i) Prove that $K = F(\omega)$.

(ii) Take any $\tau \in \mathcal{G}(K/F)$ Prove that $\tau(\omega) = \omega^k$ for some $k \in \mathbb{N}$ with $gcd(k, n) = 1$.

(iii) Prove that there is a well-defined injective group homomorphism $\mathcal{G}(K/F) \to \mathbb{Z}_n^*$ given by $\tau \mapsto [k]_n$, where $\tau(\omega) = \omega^k$. (Hence, $\mathcal{G}(K/F)$ is abelian and $|\mathcal{G}(K/F)| \,|\mathbb{Z}_n^*| = \varphi(n)$.)

5.76. Suppose that $char(F) = p \neq 0$. Take α in some extension field K of F with α algebraic over F. Prove that α is separable over F iff $\alpha \in F(\alpha^p)$.

5.77. Suppose that $char(F) = p \neq 0$. Let S and I be finite-degree field extensions of F, each lying in some field M. Suppose that S is separable over F and I is purely inseparable over F. Prove that S and I are linearly disjoint over F.

5.78. Let $F \subseteq I \subseteq K$ be fields with $[K{:}F] < \infty$ such that I is purely inseparable over F and K is normal over I. Prove that K is normal over F.

5.79. Let $F \subseteq K$ be fields with $char(F) = p \neq 0$ and $[K{:}F] < \infty$. Let S be the separable closure of F in K and I the purely inseparable closure of F in K. We know (see problem 5.71(iv)) that K is purely inseparable over S. However, this problem illustrates that K need not be separable over I. Let L be a field with $char(L) = p > 2$, let $K = L(s, t)$ where s and t are algebraically independent over L, and let $F = L(s, t^{2p} + st^p) \subseteq K$.

(i) Prove that $[K{:}F] = 2p$ (recall problem 5.25), and determine $m_{F,t}$.

(ii) Let $S = F(t^p) = L(s, t^p)$. Prove that $[S{:}F] = 2$ and that S is the separable closure of F in K.

(iii) Prove that if M is a field with $F \subsetneqq M \subsetneqq K$, then $M = S$. (Hint: $m_{M,t} | m_{F,t}$ in $K[X]$.) Hence, for the purely inseparable closure I of F in K, we have $I = F$ and K is not separable over I.

Let F be a field with $char(F) = 2$, and let C be an algebraically closed field containing F. For $\alpha \in C$, let

$$\wp(\alpha) = \alpha^2 - \alpha \tag{5.18}$$

and

$$\begin{aligned} \wp^{-1}(\alpha) &= \{\beta \in C \mid \wp(\beta) = \alpha\} \\ &= \{\text{roots of } X^2 - X - \alpha \text{ in } C\} \end{aligned} \tag{5.19}$$

Note that if $\beta \in \wp^{-1}(\alpha)$, then $\wp^{-1}(\alpha) = \{\beta, \beta + 1\}$. Let

$$\wp(F) = \{\wp(c) \mid c \in F\},$$

which is an additive subgroup of F. The next two problems show how we can classify separable quadratic extension fields of F using the function $\wp$.

5.80. Suppose that $char(F) = 2$, and let C be an algebraically closed field containing F. Let

$$f = aX^2 + bX + c \in F[X] \quad \text{with } a \neq 0.$$

(i) Suppose first that $b = 0$. Prove that f has a single root β in C, and $\beta^2 = c/a$. Prove further that f is irreducible in $F[X]$ iff $c/a \notin F^2$. Moreover, when f is irreducible, $F(\beta) = F(\sqrt{c/a}\,)$, with $F(\beta)$ purely inseparable of degree 2 over F.

(ii) Now suppose that $b \neq 0$, so that f is separable. Prove that the roots of f in C are $\beta_1 = ab^{-1}\gamma_1$ and $\beta_2 = ab^{-1}\gamma_2$, where $\{\gamma_1, \gamma_2\} = \wp^{-1}(ac/b^2)$, so $\gamma_2 = \gamma_1 + 1$. Thus, f is irreducible in $F[X]$ iff $ac/b^2 \notin \wp(F)$. When this occurs, $F(\beta_1) = F(\beta_2)$ is separable of degree 2 over F.

5.81. Let F be a field with $char(F) = 2$, and let C be an algebraically closed field containing F. Let K be a field with $F \subseteq K \subseteq C$ and $[K{:}F] = 2$. We now classify all such fields K.

(i) Suppose that K is not separable over F. Prove that for some $a \in F \setminus F^2$, we have $K = F(\sqrt{a}\,)$. Thus, K is purely inseparable over F. Note also that F^2 is a subfield of F and $K^2 = F^2 + aF^2$, which is a 2-dimensional subspace of the F^2-vector space F.

(ii) Let $K = F(\sqrt{a}\,)$ and $L = F(\sqrt{b}\,)$ in C, where $a, b \in F \setminus F^2$. Prove that K and L are F-isomorphic iff $K = L$, and that this occurs iff $b \in F^2 + aF^2$, iff $F^2 + bF^2 = F^2 + aF^2$. Thus, the purely inseparable quadratic extensions of F in C are in one-to-one correspondence with the one-dimensional F^2-subspaces of F/F^2.

(iii) Suppose that K is separable over F. Prove that $K = F(\gamma)$ for some $\gamma \in K$ with $\wp(\gamma) \in F \setminus \wp(F)$. If $c = \wp(\gamma)$, then $\wp^{-1}(c) = \{\gamma, \gamma + 1\}$ and we write $K = F(\wp^{-1}(c))$.

(iv) Let $K = F(\wp^{-1}(c))$ and $L = F(\wp^{-1}(d))$ for $c, d \in F \setminus \wp(F)$. Prove that K and L are F-isomorphic iff $K = L$, and that this occurs iff $c - d \in \wp(F)$. Thus, the separable quadratic extensions of F in C are in one-to-one correspondence with nonidentity elements of the additive group $F/\wp(F)$.

Note 5.82. *Primitive elements.* Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. An element γ of K is said to be be a *primitive element for K over F* if $K = F(\gamma)$. Recall *Steinitz' version of the Theorem of the Primitive Element*: There is a primitive element for K over F iff there are only finitely many different fields L with $F \subseteq L \subseteq K$. (Proof sketch: Suppose that $K = F(\gamma)$ and L is a field with $F \subseteq L \subseteq K$. Let

$$m_{L,\gamma} = X^n + c_{n-1}X^{n-1} + \ldots + c_1X + c_0$$

and let $M = F(c_0, \ldots, c_{n-1}) \subseteq L$. Then, $m_{M,\gamma} = m_{L,\gamma}$, hence $L = M$. Since $m_{L,\gamma}$ is a monic divisor of $m_{F,\gamma}$ in $K[X]$ there are only finitely many possibilities for $m_{L,\gamma}$. As L is determined by $m_{L,\gamma}$, there are only finitely many possible L. Conversely, suppose that there are only finitely many intermediate fields L. If $|F| < \infty$, then

$K = F(\gamma)$ for any generator γ of the finite cyclic group K^*. If F is infinite, say $K = F(\alpha_1, \ldots, \alpha_n)$, then there must be distinct $c, d \in F$ with $F(\alpha_1 + c\alpha_2) = F(\alpha_1 + d\alpha_2)$. Then, $F(\alpha_1, \alpha_2) = F(\alpha_1 + c\alpha_2)$. Hence, $K = F(\gamma)$ for some γ by induction on n.) See, e.g., Dummit & Foote [**5**, Prop. 24, p. 594] for a more detailed proof of this theorem.

5.83. Let L be a field with $char(L) = p \neq 0$. Let $K = L(s, t)$, with s and t algebraically independent over L. Let $F = L(s^p, t^p) \subseteq K$.

(i) Prove that $[K{:}F] = p^2$ and that $K^p \subseteq F$. Hence, K is purely inseparable over F.

(ii) Prove that there is no $\gamma \in K$ with $K = F(\gamma)$.

(iii) Exhibit infinitely many different fields M with $F \subseteq M \subseteq K$.

Fixed field. Let K be a field, and let S be a set of (ring) automorphisms of K. The *fixed field of* S is

$$\mathcal{F}(S) = \{\alpha \in K \mid \sigma(\alpha) = \alpha \text{ for every } \sigma \in S\}. \tag{5.20}$$

Note that $\mathcal{F}(S)$ is a subfield of K, and $S \subseteq \mathcal{G}(K/\mathcal{F}(S))$. Also, for every subfield F of K, we have the inclusions of fields,

$$F \subseteq \mathcal{F}(\mathcal{G}(K/F)) \subseteq K. \tag{5.21}$$

5.84. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. Let H be a finite subgroup of $\mathcal{G}(K/F)$, and let $L = \mathcal{F}(H)$.

(i) Take any $\alpha \in K$. Let $\{\tau(\alpha) \mid \tau \in H\} = \{\alpha_1, \ldots, \alpha_k\}$ with the α_i distinct, and let

$$f = (X - \alpha_1) \ldots (X - \alpha_i) \ldots (X - \alpha_k).$$

Prove that $f \in L[X]$ and $f = m_{L,\alpha}$.

(ii) Deduce that K is normal and separable over L.

5.7. Finite fields

We first recall basic facts about finite fields.

Existence and uniqueness of finite fields. Let F be a field with $|F| = q < \infty$. Since $char(F) \neq 0$ as F is finite, we must have $char(F) = p$ for some prime number p; so, the prime subring $P = P_F$

of F is isomorphic to $\mathbb{Z}_p$ (see (3.28)). Thus, P is a subfield of F; let $n = [F{:}P] < \infty$. Then,

$$q \;=\; |F| \;=\; |P|^{[F:P]} \;=\; p^n.$$

Since the multiplicative group F^* has order $p^n{-}1$, we have $\alpha^{p^n-1} = 1$, for each $\alpha \in F^*$. Hence,

$$\alpha^{p^n} = \alpha, \quad \text{for each } \alpha \in F.$$

Since F thus contains p^n different roots of $X^{p^n} - X \in P[X]$ (and no proper subfield is large enough to contain that many roots), F is a splitting field of $X^{p^n} - X$ over P. Hence (invoking Note 5.53), F is uniquely determined up to isomorphism by $|F| = q$, so we denote this field by $\mathbb{F}_q$.

Now take any prime number p and any $n \in \mathbb{N}$, and let

$$f \;=\; X^{p^n} - X \;\in\; \mathbb{Z}_p[X].$$

Let K be a splitting field of f over $\mathbb{Z}_p$, and let

$$K_0 \;=\; \{\text{roots of } f \text{ in } K\} \;=\; \{\alpha \in K \mid \alpha^{p^n} = \alpha\}.$$

Then, K_0 is a subfield of K, as $\mathit{char}(K) = p$. Hence, $K = K_0$, as f splits over K_0. Hence, $|K| = |K_0| = p^n$, as f is separable by the Derivative Test. Thus, for each prime power in $\mathbb{N}$ there is a unique up to isomorphism finite field with cardinality that prime power, and these are all the finite fields.

The multiplicative group of a finite field. For any finite field $\mathbb{F}_q$, its multiplicative group $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$ is a finite group of order $q-1$, and $\mathbb{F}_q^*$ is a cyclic group by problem 3.32.

Subfields of a finite field. Let L be a subfield of the finite field $\mathbb{F}_q$, where $q = p^n$ with p prime. Let $s = [\mathbb{F}_q{:}L]$. Then, $p^n = |\mathbb{F}_q| = |L|^s$. Hence, $|L| = p^d$ where $d = n/s \in \mathbb{N}$. So, $L \cong \mathbb{F}_{p^d}$, and L is the unique subfield of $\mathbb{F}_q$ of cardinality p^d, since it consists of the roots in $\mathbb{F}_q$ of $X^{p^d} - X$. Moreover, for any $e \in \mathbb{N}$ with $e|n$ the field $\mathbb{F}_q$ contains a (unique) copy of $\mathbb{F}_{p^e}$. For as $e|n$, we have $(p^e-1)|(p^n-1)$, hence $(X^{p^e-1}-1)\big|(X^{p^n-1}-1)$ in $P[X]$, where P is the prime subfield of $\mathbb{F}_q$. Therefore, $X^{p^e} - X$ splits over $\mathbb{F}_q$; the set of roots of this polynomial in $\mathbb{F}_q$ is the desired copy of $\mathbb{F}_{p^e}$. Note also that for each $m \in \mathbb{N}$, there

is a field E containing $\mathbb{F}_q$ with $[E{:}\mathbb{F}_q] = m$, namely a splitting field of $X^{q^m} - X$ over $\mathbb{F}_q$.

Galois groups for finite fields. For $q = p^n$ as above, let $P = P_{\mathbb{F}_q}$ be the prime subfield of $\mathbb{F}_q$, so $P \cong \mathbb{Z}_p$. The *Frobenius automorphism* of $\mathbb{F}_q$ is the map

$$\rho\colon \mathbb{F}_q \to \mathbb{F}_q \quad \text{given by} \quad \alpha \mapsto \alpha^p. \tag{5.22}$$

Note that ρ is a ring homomorphism as $char(\mathbb{F}_q) = p$, and ρ is injective as $\mathbb{F}_q$ is a field, hence surjective as $\mathbb{F}_q$ is finite. Hence, ρ is field automorphism of $\mathbb{F}_q$. Moreover, $\rho(\alpha) = \alpha$ iff α is a root of $X^p - X \in P[X]$. Since P contains p (hence, all) roots of this polynomial, $\mathcal{F}(\rho) = P$. Thus, $\rho \in \mathcal{G}(\mathbb{F}_q/P)$. For $i \in \mathbb{N}$, we have

$$\big|\{\alpha \in \mathbb{F}_q \mid \rho^i(\alpha) = \alpha\}\big| = \big|\{\text{roots of } X^{p^i} - X \text{ in } \mathbb{F}_q\}\big| \le p^i; \tag{5.23}$$

hence, $\rho^i \neq id_{\mathbb{F}_q}$ for $1 \le i < n$, but $\rho^n = id_{\mathbb{F}_q}$. Thus, $|\rho| = n$ in $\mathcal{G}(\mathbb{F}_q/P)$. Since $|\,\mathcal{G}(\mathbb{F}_q/P)| \le [\mathbb{F}_q{:}P] = n$ (see (5.15)), it follows that $\mathcal{G}(\mathbb{F}_q/P)$ is the cyclic group generated by ρ, of order n. Now let L be any subfield of $\mathbb{F}_q$. As we have seen, $L = \mathbb{F}_{p^d}$, for some d dividing n. Equation (5.23) shows that $\rho^i|_L \neq id_L$ for $1 \le i \le d-1$, but $\rho^d|_L = id_L$. Since $\mathcal{G}(\mathbb{F}_q/L)$ is a subgroup of $\mathcal{G}(\mathbb{F}_q/P) = \langle\rho\rangle$ it follows that $\mathcal{G}(\mathbb{F}_q/L) = \langle\rho^d\rangle$. Hence,

$$\mathcal{F}(\mathcal{G}(\mathbb{F}_{p^n}/\mathbb{F}_{p^d})) = \mathcal{F}(\rho^d) = \mathbb{F}_{p^d} \tag{5.24}$$

and

$$|\,\mathcal{G}(\mathbb{F}_{p^n}/\mathbb{F}_{p^d})| = |\rho^d| = n/d = [\mathbb{F}_{p^n}{:}\mathbb{F}_{p^d}]. \tag{5.25}$$

5.85. Let r be a prime number. Determine how many irreducible polynomials there are of degree r in $\mathbb{F}_q[X]$. The prime r may or may not divide q.

5.86. Let $f = X^n - 1 \in \mathbb{F}_\ell[X]$, where $\ell = p^k$ with p prime and $p \nmid n$. Let E be a splitting field of f over $\mathbb{F}_\ell$. Recall (see problem 5.75) that the set U of roots of f in E is a cyclic group with $|U| = n$.

(i) Prove that f splits over a finite field $M \supseteq \mathbb{F}_\ell$ iff $n \,\big|\, |M^*|$.

(ii) Deduce that $[E{:}\mathbb{F}_\ell]$ is the least integer r such that $n \,|\, (\ell^r - 1)$. That is, r is the order of the element $[\ell]_n$ in the multiplicative group $\mathbb{Z}_n^*$.

(iii) Now, suppose that n is a prime number. Then, U consists of 1 together with the $n-1$ primitive n-th roots of unity. Prove that the irreducible factorization of f in $\mathbb{F}_\ell$ has the form

$$f = (X-1)g_1 g_2 \dots g_s,$$

where the g_i are distinct monic irreducibles in $\mathbb{F}_\ell[X]$, and each g_i has degree r (the r of part (ii)); so, $s = (n-1)/r$.

(iv) Still assume that n is prime, as in part (iii). Since E and $\mathbb{F}_\ell$ are finite fields, we know that $\mathcal{G}(E/\mathbb{F}_\ell)$ is a finite cyclic group of order $[E{:}\mathbb{F}_\ell] = r$, and $\mathcal{F}(\mathcal{G}(E/\mathbb{F}_\ell)) = \mathbb{F}_\ell$ (see (5.24) and (5.25)). Say $\mathcal{G}(E/\mathbb{F}_\ell) = \langle\tau\rangle$. Then, τ maps U bijectively to itself; let σ be a permutation in the symmetric group S_n corresponding to the action of τ on U. That is, let $U = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ and define $\sigma \in S_n$ by $\tau(\alpha_i) = \alpha_{\sigma(i)}$ for all i. Prove that in its disjoint cycle decomposition σ is a product of s cycles with each cycle of length r. (Here r and s are as in part (iii). There is one cycle for each irreducible factor g_i of f.)

5.87. Let $\mathbb{F}_q$ be a finite field, and let $\alpha \in \mathbb{F}_q^*$. Let K be a splitting field over $\mathbb{F}_q$ of $X^{q+1} - \alpha$. Prove that $[K{:}\mathbb{F}_q] = 2$ (so $K \cong \mathbb{F}_{q^2}$).

5.88. Prove that $X^4 + 1$ in $\mathbb{Z}[X]$ is irreducible in $\mathbb{Q}[X]$, but that its image in $\mathbb{Z}_p[X]$ is reducible for every prime p. (Recall problem 5.51.)

5.89. Let $\mathbb{F}$ be a finite field. Prove that every element of $\mathbb{F}$ is expressible as $\alpha^2 + \beta^2$ for some $\alpha, \beta \in \mathbb{F}$.

5.90. Let $\mathbb{F}_q$ be a finite field and let A be an algebraic closure of $\mathbb{F}_q$. Prove that $\mathcal{G}(A/\mathbb{F}_q)$ is abelian and uncountable. (Hint: Apply the generalized Isomorphism Extension Theorem (problem 5.55) to see that every automorphism of a subfield of A extends to an automorphism of A.) Prove also that every nonidentity element of $\mathcal{G}(A/\mathbb{F}_q)$ has infinite order.

5.8. Galois field extensions

Galois field extensions. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. The field K is said to be *Galois over* F if $\mathcal{F}(\mathcal{G}(K/F)) = F$.

5.91. *Galois connections.* The notion of a Galois connection captures the elementary formalism of the Galois correspondence between field extensions and Galois groups. Let S and T be two partially ordered sets, and let $f\colon S \to T$ and $g\colon T \to S$ be functions. We say that f and g give a *Galois connection* between S and T if the following hold for all $s, s_1, s_2 \in S$ and $t, t_1, t_2 \in T$:

(i) if $s_1 \leq s_2$, then $f(s_2) \leq f(s_1)$;

(ii) if $t_1 \leq t_2$, then $g(t_2) \leq g(t_1)$;

(iii) $s \leq g(f(s))$;

(iv) $t \leq f(g(t))$.

Prove that when this occurs, $f(g(f(s))) = f(s)$ for all $s \in S$ and, likewise, $g(f(g(t))) = g(t)$ for all $t \in T$. Deduce that f and g induce a one-to-one order-reversing correspondence between $im(g)$ and $im(f)$.

Example 5.92. Let $F \subseteq K$ be fields, and let $G = \mathcal{G}(K/F)$. Let S be the set of subgroups H of G, and let T be the set of fields L with $F \subseteq L \subseteq K$, with the partial orderings on S and T given by inclusion of sets. Let $f\colon S \to T$ be given by $f(H) = \mathcal{F}(H)$ and $g\colon T \to S$ be given by $g(L) = \mathcal{G}(K/L)$. It is easy to check that f and g give a Galois connection between S and T. Thus, the preceding problem shows that there is a one-to-one correspondence (given by f and g) between (i) those subgroups H of G of the form $\mathcal{G}(K/L)$ for some field L with $F \subseteq L \subseteq K$; and (ii) the fields $\mathcal{F}(H)$ for subgroups H of G. In general, it can be difficult to tell which subgroups and which intermediate fields are part of this one-to-one correspondence. But when K is a Galois extension of F, part (i) of the Fundamental Theorem (see p. 228) says that the maps f and g are surjective; hence *all* the subgroups and *all* the intermediate fields are then included in the one-to-one correspondence.

Recall the *Characterization Theorem* for Galois extensions: Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. Then the following conditions are equivalent:

(a) K is Galois over F, i.e., $F = \mathcal{F}(\mathcal{G}(K/F))$.

(b) $|\,\mathcal{G}(K/F)| = [K{:}F]$.

(c) K is normal and separable over F.

(d) K is a splitting field over F of some separable polynomial $f \in F[X]$.

See, e.g., Dummit & Foote [**5**, pp. 562–574] or Cox [**4**, Th. 7.1.1, p. 147; Th. 7.1.5(c), p. 150] for proofs of the Characterization Theorem.

5.93. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. Let H be a subgroup of the finite group $\mathcal{G}(K/F)$, and let $L = \mathcal{F}(H)$. The goal of this problem is to show that $H = \mathcal{G}(K/L)$. Note that K is normal and separable over L by problem 5.84(ii), so K is Galois over L by the Characterization Theorem.

(i) Let M be a field with $L \subseteq M \subseteq K$. Then, K is Galois over M by the Characterization Theorem (c). Deduce that there are only finitely many such fields M, as there are only finitely many subgroups of $\mathcal{G}(K/L)$.

(ii) The Theorem of the Primitive Element (see Note 5.82) and part (i) show that $K = L(\gamma)$ for some $\gamma \in K$. Prove that

$$|H| = deg(m_{L,\gamma}) = [K{:}L].$$

Deduce that $H = \mathcal{G}(K/L)$.

Recall the *Fundamental Theorem of Galois Theory*: Let $F \subseteq K$ be fields with K Galois over F, and let $G = \mathcal{G}(K/F)$. Then,

(i) There is a one-to-one inclusion-reversing correspondence between (all) the fields L with $F \subseteq L \subseteq K$ and (all) the subgroups H of G such that when L corresponds to H,

$$H = \mathcal{G}(K/L) \quad \text{and} \quad L = \mathcal{F}(H).$$

(ii) When L corresponds to H, the field K is Galois over L, and

$$|H| = [K{:}L] \qquad \text{and} \qquad |G{:}H| = [L{:}F].$$

(iii) When L corresponds to H, the field L is Galois over F iff H is normal in G; when this occurs,

$$\mathcal{G}(L/F) \cong G/H.$$

See, e.g., Dummit & Foote [**5**, Th. 14, p. 574] or Cox [**4**, Th. 7.3.1, Th. 7.3.2, pp. 162–163] or Hungerford [**9**, Th. 2.5, p. 245] for proofs of

the Fundamental Theorem for fields of any characteristic. Note that part (i) of the Fundamental Theorem follows from the Characterization Theorem (c) and problem 5.93; they show that all intermediate fields L and all subgroups of $\mathcal{G}(K/F)$ are included in the correspondence of the Galois connection described in Example 5.92.

5.94. Let $F \subseteq K$ be fields with K Galois over F, and let $G = \mathcal{G}(K/F)$.

(i) Let L be any field with $F \subseteq L \subseteq K$, and let $H = \mathcal{G}(K/L)$, a subgroup of G. Take any $\sigma \in G$, and let $L' = \sigma(L)$, which is a subfield of K isomorphic to L. Prove that

$$\mathcal{G}(K/L') \;=\; \sigma H \sigma^{-1}.$$

(ii) Let L_1, L_2 be fields with $F \subseteq L_i \subseteq K$ for $i = 1, 2$. Let $H_i = \mathcal{G}(K/L_i) \subseteq G$. Prove that L_2 is F-isomorphic to L_1 iff there is $\tau \in G$ such that $H_2 = \tau H_1 \tau^{-1}$.

Example 5.95. Let $F = \mathbb{Q}$, $K = \mathbb{Q}(\sqrt{2}\,, \sqrt{3}\,)$, $L_1 = \mathbb{Q}(\sqrt{2}\,)$, and $L_2 = \mathbb{Q}(\sqrt{3}\,)$. Then, K is Galois over F with $\mathcal{G}(K/F) \cong C_2 \times C_2$. The groups (of order 2) $\mathcal{G}(K/L_1)$ and $\mathcal{G}(K/L_2)$ are isomorphic but not conjugate in the abelian group $\mathcal{G}(K/F)$, and $L_1 \ncong L_2$.

5.96. Let $F \subseteq L \subseteq K$ be fields with K Galois over F; let $G = \mathcal{G}(K/F)$ and $H = \mathcal{G}(K/L)$. This problem shows how $\mathcal{G}(L/F)$ fits into the picture when L is not Galois over F. For this, let $M = \mathcal{F}(\mathcal{G}(L/F))$ which is a field with $F \subseteq M \subseteq L$. Let $N = \mathcal{G}(K/M)$, which is a subgroup of G with $H \subseteq N$.

(i) Prove that $N = N_G(H)$, the normalizer of H in G.

(ii) Prove that $\mathcal{G}(L/F) = \mathcal{G}(L/M) \cong N/H$.

5.97. Let $F \subseteq L$ be fields with $[L{:}F] < \infty$, and suppose there is $\alpha \in L$ with $L = F(\alpha)$.

(i) Prove that the number of roots of $m_{F,\alpha}$ in L equals $|\mathcal{G}(L/F)|$. Thus, this number depends only on L and F, and not on the choice of α.

(ii) Deduce that the number of roots of $m_{F,\alpha}$ in L divides $[L{:}F]$.

5.98. Suppose that $char(F) \neq 2$, let

$$f \;=\; X^4 + 2bX^2 + c$$

be irreducible in $F[X]$, and let E be a splitting field of f over F. Then, E is Galois over F, as f is separable. Let $G = \mathcal{G}(E/F)$. Assume that $[E{:}F] = 8$. We have seen in problem 5.63 that $G \cong D_4$. More specifically, let $\pm\alpha, \pm\beta$ be the distinct roots of f in E; let $\sigma \in G$ be the F-automorphism of E such that $\sigma(\alpha) = \beta$ and $\sigma(\beta) = -\alpha$; and let $\tau \in G$ be the automorphism such that $\tau(\alpha) = \alpha$ and $\tau(\beta) = -\beta$. Then $|\sigma| = 4$, $|\tau| = 2$, $\tau \notin \langle\sigma\rangle$, $\tau\sigma\tau^{-1} = \sigma^{-1}$, and $G = \langle\sigma\rangle \cup \tau\langle\sigma\rangle$. Note that G has five subgroups of order 2 and three subgroups of order 4.

(i) For each subgroup H of G, determine the fixed field $\mathcal{F}(H)$, (preferably in the form $\mathcal{F}(H) = F(\gamma)$ for some $\gamma \in E$), and verify that $[E{:}\mathcal{F}(H)] = |H|$. (Here is one way of finding elements of $\mathcal{F}(H)$: For any $\delta \in E$ note that $\prod_{\rho\in H} \rho(\delta) \in \mathcal{F}(H)$ and $\sum_{\rho\in H} \rho(\delta) \in \mathcal{F}(H)$.)

(ii) For each normal subgroup H of G, verify that the field $\mathcal{F}(H)$ is normal over F by finding a polynomial in $F[X]$ for which $\mathcal{F}(H)$ is a splitting field over F.

(iii) For each nonnormal subgroup H of G, prove that its fixed field $\mathcal{F}(H)$ is not normal over F by finding an irreducible polynomial in $F[X]$ that has a root in $\mathcal{F}(H)$ but does not split over $\mathcal{F}(H)$.

5.99. For each $n \in \mathbb{N}$, give an example of fields $F \subseteq K$ with $[K{:}F] = n$ such that there are no fields L with $F \subsetneqq L \subsetneqq K$.

The next two problems give Artin's proof of the Fundamental Theorem of Algebra that the field $\mathbb{C}$ is algebraically closed. The proof uses the following two consequences of the Intermediate Value Theorem (IVT) for continuous functions $\mathbb{R} \to \mathbb{R}$:

(i) Every polynomial in $\mathbb{R}[X]$ of odd degree has a root in $\mathbb{R}$.

(ii) Every positive real number has a square root in $\mathbb{R}$. (For $c \in \mathbb{R}$ with $c > 0$, apply the IVT to $X^2 - c \in \mathbb{R}[X]$.)

5.100. Using the facts just quoted, prove the following:

(i) There is no finite-degree field extension K of $\mathbb{R}$ with $[K{:}\mathbb{R}]$ an odd integer, except $K = \mathbb{R}$.

(ii) There is no field extension L of $\mathbb{C}$ with $[L{:}\mathbb{C}] = 2$.

5.101. Now prove the Fundamental Theorem of Algebra as follows, using the results of the preceding problem: For purposes of contradiction, suppose there is a field $K \supseteq \mathbb{C}$ with $1 < [K{:}\mathbb{C}] < \infty$. Let E be a normal closure of K over $\mathbb{R}$. Then $E \supseteq K \supsetneq \mathbb{C}$, $[E{:}\mathbb{R}] < \infty$, and E is Galois over $\mathbb{R}$, since it is normal and separable over $\mathbb{R}$. (The separability is free since $char(\mathbb{R}) = 0$.) Let $G = \mathcal{G}(E/\mathbb{R})$.

(i) Let P be a 2-Sylow subgroup of G, and let $L = \mathcal{F}(P)$. Prove that $[L{:}\mathbb{R}]$ is odd. Deduce that $L = \mathbb{R}$ and $P = G$, i.e., G is a 2-group.

(ii) Let $H = \mathcal{G}(E/\mathbb{C}) \subseteq G$. Prove that H has a subgroup H_0 with $|H{:}H_0| = 2$, and let $M = \mathcal{F}(H_0)$. Prove that $[M{:}\mathbb{C}] = 2$. This contradicts the preceding problem. Hence, the postulated field K cannot exist, proving that $\mathbb{C}$ is algebraically closed.

5.102. Let field $K = L(t_1, \ldots, t_n)$, where L is a field and $t_1, \ldots, t_n$ are algebraically independent over L. Let

$$\begin{aligned} f &= (X - t_1)(X - t_2)\ldots(X - t_n) \\ &= X^n - s_1X^{n-1} + \ldots + (-1)^{n-j}s_{n-j}X^j + \ldots + (-1)^n s_n \in K[X], \end{aligned}$$

where

$$\begin{aligned} s_1 &= t_1 + \ldots + t_n, \quad \ldots, \\ s_j &= \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} t_{i_1}t_{i_2}\ldots t_{i_j}, \quad \ldots, \\ s_n &= t_1t_2\ldots t_n. \end{aligned} \tag{5.26}$$

The s_i are called the *elementary symmetric polynomials in the* t_i. Let $F = L(s_1, s_2, \ldots, s_n) \subseteq K$. Thus, $f \in F[X]$, and K is a splitting field of f over F. Hence, $[K{:}F] \le n!$ and K is Galois over F.

(i) Prove that $s_1, s_2, \ldots, s_n$ are algebraically independent over L.

(ii) Every permutation $\sigma \in S_n$ induces an L-automorphism π_σ of K given by $t_i \mapsto t_{\sigma(i)}$ for each i. Thus, for any $g, h \in L[t_1, t_2, \ldots, t_n]$ with $h \neq 0$,

$$\pi_\sigma(g/h) = g(t_{\sigma(1)}, \ldots, t_{\sigma(n)})/h(t_{\sigma(1)}, \ldots, t_{\sigma(n)}).$$

Then, $\pi_\sigma(s_j) = s_j$ for every j; hence, $\pi_\sigma \in \mathcal{G}(K/F)$. The map $\psi\colon S_n \to \mathcal{G}(K/F)$ given by $\sigma \mapsto \pi_\sigma$ is clearly an injective group homomorphism. Prove that ψ is an isomorphism. Hence, $\mathcal{G}(K/F) \cong S_n$ and $[K{:}F] = n!$.

Note: This shows that every element of $L(t_1, \ldots, t_n)$ invariant under all permutations of the t_i (i.e., mapped to itself by all the π_σ) is a rational function of the s_i, meaning that it lies in $L(s_1, \ldots, s_n)$. A stronger result is known: $F[t_1, \ldots, t_n] \cap L(s_1, \ldots, s_n) = L[s_1, \ldots, s_n]$. That is, every symmetric polynomial in the t_i is expressible as a polynomial in the elementary symmetric polynomials of the t_i. See, e.g., Cox [**4**, Th. 2.2.2, p. 30] for a proof.

5.103. Let

$$K = \mathbb{Q}\big(\sqrt{(5+\sqrt{5})(21+\sqrt{21}\,)}\,\big).$$

Prove that $[K{:}\mathbb{Q}] = 8$ and that K is Galois over $\mathbb{Q}$ with $\mathcal{G}(K/\mathbb{Q})$ the quaternion group of order 8.

5.104. Let L and K be extension fields of F, each lying in some field M. Suppose that K is Galois over F.

(i) Prove that the compositum $L{\cdot}K$ is Galois over L.

(ii) Prove that there is a well-defined group homomorphism $\theta\colon \mathcal{G}(L{\cdot}K\,/\,L) \to \mathcal{G}(K/F)$ given by $\tau \mapsto \tau|_K$.

(iii) Prove that θ is injective and that $\mathcal{F}(im(\theta)) = K \cap L$.

(iv) Deduce that

$$\mathcal{G}(L{\cdot}K\,/\,L) \;\cong\; \mathcal{G}(K/(K \cap L)), \tag{5.27}$$

and that $[L{\cdot}K : L] = [K{:}(K \cap L)]$. (Thus, K and L are linearly disjoint over $K \cap L$.) The isomorphism of (5.27) is called the Theorem on Natural Irrationalities. See Cox [**4**, pp. 337–339] for an explanation of this theorem name.

5.105. Let L and K be Galois extensions of a field F, with L and K each lying in some field M.

(i) Prove that the compositum $L{\cdot}K$ is Galois over F.

(ii) Prove that

$$\mathcal{G}(L{\cdot}K/F) \;\cong\; \mathcal{G}(L/F) \times \mathcal{G}(K/F) \;\text{ iff }\; L \cap K \;=\; F.$$

5.106. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. If K is separable over F, then a normal closure E of K over F is Galois over F by problem 5.73 and the Characterization Theorem (c) (see p. 227). So, there are only finitely many fields L with $F \subseteq L \subseteq K$, since there are only finitely many intermediate fields between F and E by the Galois correspondence. So, by the Theorem of the Primitive Element (see Note 5.82), $K = F(\gamma)$ for some γ. In particular, whenever $char(F) = 0$, K is separable over F, so $K = F(\gamma)$. Now, suppose that $char(F) = p \neq 0$ and $K = F(\alpha_1, \ldots, \alpha_n, \beta)$ with each α_i separable over F and β algebraic over F. Prove that $K = F(\delta)$, for some $\delta \in K$.

5.107. When $char(F) = 0$, normal field extensions of F are the same as Galois extensions. Now assume that $char(F) = p \neq 0$. This problem shows how normal and Galois extensions of F are related. Let E be a finite-degree extension field of F, and suppose that E is normal over F. Let $I = \mathcal{F}(\mathcal{G}(E/F))$, and let S be the separable closure of F in E (as in problem 5.71). So, E is purely inseparable over S.

(i) Prove that I is purely inseparable over F and that E is Galois over I, with $\mathcal{G}(E/I) = \mathcal{G}(E/F)$.

(ii) Prove that S is Galois over F, with $\mathcal{G}(E/F) \cong \mathcal{G}(S/F)$ via the map $\tau \mapsto \tau|_S$.

(iii) Prove that $S \cap I = F$, $E = S{\cdot}I$, and that $[E{:}I] = [S{:}F]$.

Note that the converse to the preceding problem holds: If $char(F) = p \neq 0$ and I and S are finite-degree field extensions of F lying in a common field M, with I purely inseparable over F and S Galois over F, then $E = S{\cdot}I$ is normal over F with S the separable closure of F in E and $I = \mathcal{F}(\mathcal{G}(E/F))$.

5.108. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$, and let I be the purely inseparable closure of F in K. Prove that K is separable over I iff there is a finite-degree separable field extension L of K with L normal over F.

5.109. Let $\mathbb{F}_q$ be the finite field with q elements, and let t be transcendental over $\mathbb{F}_q$. Recall from (5.16) that

$$\mathcal{G}(\mathbb{F}_q(t)/\mathbb{F}_q) \cong PGL_2(\mathbb{F}_q),$$

so by (2.61),

$$\big|\,\mathcal{G}(\mathbb{F}_q(t)/\mathbb{F}_q)\big| = (q^2-1)(q^2-q)/(q-1) = q(q^2-1).$$

Prove that

$$\mathcal{F}(\mathcal{G}(\mathbb{F}_q(t)/\mathbb{F}_q)) = \mathbb{F}_q(s), \quad \text{where} \quad s = (t^{q^2}-t)^{q+1}/(t^q-t)^{q^2+1}.$$

(By contrast, if F is an infinite field, then $\big|\mathcal{G}(F(t)/F)\big| = \infty$. It follows by problem 5.25(iv) and (5.15) that $\mathcal{F}(\mathcal{G}(F(t)/F)) = F$.) (Hint: Use problem 5.93.)

5.9. Cyclotomic polynomials and cyclotomic extensions

Cyclotomic polynomials. Fix $n \in \mathbb{N}$. The polynomial $X^n - 1$ in $\mathbb{C}[X]$ splits over $\mathbb{C}$ with roots $e^{2\pi ij/n}$ for $j = 0, 1, \ldots, n-1$, which are the n-th roots of unity in $\mathbb{C}$. These roots make up a cyclic group of order n. The primitive n-th roots of unity are the elements of order n in this group, which are the elements that generate the cyclic group; there are $\varphi(n)$ of them, where φ is Euler's φ-function (see (1.10) and (1.11) above). The *n-th cyclotomic polynomial*, denoted Ψ_n, is the monic separable polynomial in $\mathbb{C}[X]$ with roots all the primitive n-th roots of unity. Thus,

$$\Psi_n = \prod_{j=1}^{\varphi(n)} (X - \omega_j) \tag{5.28}$$

where $\{\omega_1, \ldots, \omega_{\varphi(n)}\} = \{e^{2\pi ik/n} \mid k \in \mathbb{Z},\ gcd(k,n) = 1\}$. We recall in (5.29)–(5.31) three key properties of cyclotomic polynomials:

$$X^n - 1 = \prod_{\substack{d|n \\ 1\le d\le n}} \Psi_d. \tag{5.29}$$

This holds because the monic separable polynomials on each side of the equality have as their roots all the n-th roots of unity. (For $d|n$, the primitive d-th roots of unity are the n-th roots of unity of order d in $\mathbb{C}^*$.)

$$\Psi_n \in \mathbb{Z}[X]. \tag{5.30}$$

This follows by induction from (5.29): For $n > 1$, let $g = \prod_{\substack{d|n \\ 1 \le d < n}} \Psi_d$. By induction, $g \in \mathbb{Z}[X]$. Since g is monic, by the Division Algorithm (problem 3.26 above) we can divide g into $X^n - 1$ in $\mathbb{Z}[X]$, and the quotient Ψ_n lies in $\mathbb{Z}[X]$.

$$\Psi_n \text{ is irreducible in } \mathbb{Z}[X], \text{ so also irreducible in } \mathbb{Q}[X]. \tag{5.31}$$

See any text covering Galois theory for a proof of (5.31).

5.110. Let p be a prime number, and let $n \in \mathbb{N}$. Prove without invoking (5.31) above that Ψ_{p^n} is irreducible in $\mathbb{Q}[X]$. (Hint: Apply Eisenstein's Irreducibility Criterion to $\Psi_{p^n}(X+1)$.)

5.111. Let p be prime number, and let $n \in \mathbb{N}$.

(i) Prove that if $p|n$, then

$$\Psi_{pn}(X) = \Psi_n(X^p).$$

(ii) Prove that if $p \nmid n$, then

$$\Psi_{pn}(X) = \Psi_n(X^p) \big/ \Psi_n(X).$$

(iii) Prove that if n is odd and $n > 1$, then

$$\Psi_{2n}(X) = \Psi_n(-X).$$

Note that these formulas simplify the task of computing Ψ_n: Let $n = p_1^{r_1} \dots p_k^{r_k}$ for distinct primes p and for $r_i \in \mathbb{N}$; let $m = p_1 p_2 \dots p_k$. Then, $\Psi_n(X) = \Psi_m(X^{n/m})$ by part (i), and Ψ_m can be computed by repeated application of part (ii).

5.112. Suppose that a cyclotomic polynomial Ψ_n has some coefficient different from 0, 1, and -1. Prove that n has at least three different odd prime factors.

5.113. Let $n, k \in \mathbb{N}$, and write $k = ab$, where every prime factor of a divides n and no prime factor of b divides n. Prove that

$$\Psi_n(X^k) = \prod_{\substack{d|b \\ 1 \le d \le b}} \Psi_{nad}(X).$$

5.114. Let p be a prime number, and let $n \in \mathbb{N}$ with $p \nmid n$. Let $\overline{\Psi_n}$ be the image of Ψ_n in $\mathbb{Z}_p[X]$.

(i) Prove that the roots of $\overline{\Psi_n}$ in a splitting field are primitive n-th roots of unity. (Hint: Reduce formula (5.29) *mod p.*)

(ii) Prove that $\overline{\Psi_n}$ has a root in $\mathbb{Z}_p$ iff $n|(p-1)$.

(iii) Prove that there are infinitely many prime numbers p with $p \equiv 1 \ (mod\ n)$. (Recall problem 3.33.) This is a special case of Dirichlet's Theorem on primes in an arithmetic progression; see the comments on p. 113.

Cyclotomic extensions of $\mathbb{Q}$. For any $n \in \mathbb{N}$, let ω be a primitive n-th root of unity in $\mathbb{C}$. Let

$$\mathbb{Q}_n \ = \ \mathbb{Q}(\omega) \subseteq \mathbb{C}.$$

This $\mathbb{Q}_n$ is called the *n-th cyclotomic extension of* $\mathbb{Q}$. Note that since $m_{\mathbb{Q},\omega} = \Psi_n$, we have

$$[\mathbb{Q}_n{:}\mathbb{Q}] \ = \ deg(\Psi_n) \ = \ \varphi(n). \tag{5.32}$$

Also, $\mathbb{Q}_n$ is a splitting field of $X^n - 1$ over $\mathbb{Q}$ (since the roots of $X^n - 1$ are the powers of ω), hence $\mathbb{Q}_n$ is Galois over $\mathbb{Q}$, so

$$|\,\mathcal{G}(\mathbb{Q}_n/\mathbb{Q})| \ = \ [\mathbb{Q}_n{:}\mathbb{Q}] \ = \ \varphi(n) \ = \ |\mathbb{Z}_n^*|.$$

Hence, the injective group homomorphism $\mathcal{G}(\mathbb{Q}_n/\mathbb{Q}) \to \mathbb{Z}_n^*$ given by $\tau \mapsto [k]_n$ where $\tau(\omega) = \omega^k$ (see problem 5.75) must be an isomorphism; hence,

$$\mathcal{G}(\mathbb{Q}_n/\mathbb{Q}) \ \cong \ \mathbb{Z}_n^*. \tag{5.33}$$

5.115. For $n \in \mathbb{N}$, consider the n-th cyclotomic extension $\mathbb{Q}_n$ of $\mathbb{Q}$.

(i) Let W be the group of all roots of unity (of all orders) in $\mathbb{Q}_n$. Prove that $|W| = n$ if n is even, and $|W| = 2n$ if n is odd.

(ii) Prove that if n is odd, then $\mathbb{Q}_n = \mathbb{Q}_{2n}$.

(iii) Prove that for $n, k \in \mathbb{N}$ with k even, $\mathbb{Q}_n \subseteq \mathbb{Q}_k$ iff $n|k$.

5.116. Let $m, n \in \mathbb{N}$, and let $d = gcd(m,n)$ and $\ell = lcm(m,n)$. Prove that

$$\mathbb{Q}_m \cdot \mathbb{Q}_n \ = \ \mathbb{Q}_\ell \qquad \text{and} \qquad \mathbb{Q}_m \cap \mathbb{Q}_n \ = \ \mathbb{Q}_d.$$

5.117. Let K be finite-degree field extension of $\mathbb{Q}$. Prove that K contains only finitely many roots of unity.

5.118. Let A be any finite abelian group. Prove that there is a field $K \supseteq \mathbb{Q}$ with K Galois over $\mathbb{Q}$ and $\mathcal{G}(K/\mathbb{Q}) \cong A$. (Hint: Find a suitable K in some cyclotomic extension of $\mathbb{Q}$.)

5.119. Let $n \in \mathbb{N}$ with $n > 1$, and suppose that $char(F) \nmid n$. Choose $c \in F^*$, and let

$$f = X^n - c \in F[X].$$

Assume that f is irreducible in $F[X]$. (See Note 5.132 below for when this occurs.) Let K be splitting field of f over F. Then, K is Galois over F since the Derivative Test shows that f is separable. Let $G = \mathcal{G}(K/F)$. We have $K = F(\alpha, \omega)$ where α is a root of f and ω is a primitive n-th root of unity. Let

$$H = \mathcal{G}(K/F(\alpha)) \subseteq G \qquad \text{and} \qquad N = \mathcal{G}(K/F(\omega)) \subseteq G.$$

Since K is a splitting field of $X^n - 1$ over $F(\alpha)$ we know (see problem 5.75) that H is isomorphic to a subgroup of $\mathbb{Z}_n^*$, so H is abelian and $|H| \big| \varphi(n)$. Note that

$$|G| = [K{:}F] = [F(\alpha){:}F]\,[K{:}F(\alpha)] = n\,|H|. \tag{5.34}$$

Also, as $F(\omega)$ is Galois over F (since it is a splitting field of $X^n - 1$ over F) the Fundamental Theorem (see p. 228) says that N is normal in G, and $G/N \cong \mathcal{G}(F(\omega)/F)$. Let $k = |N|$.

(i) Prove that there is a well-defined injective group homomorphism $N \to (\mathbb{Z}_n, +)$ given by $\tau \mapsto [i]_n$ where $\tau(\alpha) = \alpha\omega^i$. Hence, N is a cyclic group and $k|n$. Prove also that $k > 1$.

(ii) Prove that $\mathcal{F}(NH) = F(\alpha) \cap F(\omega) = F(\alpha^k)$. (Note that $[F(\alpha){:}F(\alpha^k)] = k$ since $m_{F,\alpha} = f \in F[X^k]$; recall problem 5.11.)

(iii) Prove that $\mathcal{G}(K/F(\alpha^k))$ is a semidirect product of N by H. (It follows that G is a semidirect product of N by H iff $|G| = |N|\,|H|$, iff $k = n$ (see (5.34)).

(iv) Prove that the only fields L with $F(\alpha^k) \subseteq L \subseteq F(\alpha)$ are $L = F(\alpha^m)$ for some $m \in \mathbb{N}$ with $m|k$. (Hint: Translate this into a problem about groups.)

(v) Prove that $H \cap Z(G) = \{id_K\}$. (Hint: Prove that there is $\sigma \in G$ with $\sigma(\alpha) = \alpha\omega$. Prove that σ does not commute

with any nonidentity element of H.) It follows that if G is abelian, then $|H| = 1$, i.e, $F(\omega) \subseteq F(\alpha) = K$.

(vi) Prove that $F(\omega^k) \subseteq F(\alpha^k)$. (Hint: Apply part (v) with $g = X^{n/k} - c = m_{F,\alpha^k}$ replacing f.)

5.120. In the setting of the preceding problem, make the added assumption that $F = \mathbb{Q}$.

(i) Prove that if $k = n$, then $[K{:}\mathbb{Q}] = n\varphi(n)$ and

$$\mathcal{G}(K/\mathbb{Q}) \cong \mathit{Hol}(\mathbb{Z}_n) \cong \mathit{Aff}_n$$

(cf. Example 2.71(i)).

(ii) Prove that n/k is a power of 2. It follows that if n is odd, then $k = n$.

(iii) Prove that if $c > 0$ then $n/k = 1$ or $= 2$.

Example 5.121. Here are some examples illustrating the preceding two problems:

(i) For any $m \in \mathbb{N}$, let

$$f = X^{2^m} + 1 = \Psi_{2^{m+1}},$$

which is irreducible in $\mathbb{Q}[X]$ (see (5.31)). Then, $K = \mathbb{Q}_{2^{m+1}}$ is a splitting field of f over $\mathbb{Q}$, and a root α of f in $\mathbb{Q}$ is a primitive 2^{m+1}-root of unity; also, $\omega = \alpha^2$ is a primitive 2^m-th root of unity. Here, $n = \mathit{deg}(f) = 2^{m+1}$, while $k = 2$, as $\mathbb{Q}(\alpha) \cap \mathbb{Q}(\omega) = \mathbb{Q}(\alpha^2)$; so $n/k = 2^{m-1}$.

(ii) In the context of problem 5.119(iv) when $k = n$, the only fields L with $F \subseteq L \subseteq F(\alpha)$ are the obvious fields $L = F(\alpha^\ell)$ for $\ell | n$. However, when $k < n$ there can be further fields L. For example, let $F = \mathbb{Q}$ and let $f = X^6 + 3$, which is irreducible in $\mathbb{Q}[X]$ (e.g., by Eisenstein's criterion, or by problem 5.48). For $\alpha = \sqrt[6]{-3} \in \mathbb{C}$ and $\omega = (1+\sqrt{-3})/2$, a primitive 6-th root of unity, we have $\mathbb{Q}(\alpha^3) = \mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\omega)$, so $k = 3$ while $n = 6$. The distinct fields L with $\mathbb{Q} \subsetneqq L \subsetneqq \mathbb{Q}(\alpha)$ are $\mathbb{Q}(\alpha^3)$, $\mathbb{Q}(\alpha^2)$, $\mathbb{Q}(\alpha^2\omega^2)$, and $\mathbb{Q}(\alpha^2\omega^4)$. The last three of these fields are isomorphic though distinct (as $\omega^2, \omega^4 \notin \mathbb{Q}(\alpha^2)$). Note also that for the splitting field $K = \mathbb{Q}(\alpha, \omega) = \mathbb{Q}(\alpha)$

of f over $\mathbb{Q}$, we have $\mathcal{G}(K/\mathbb{Q}) \cong S_3$, which is nonabelian, even though $H = \mathcal{G}(K/\mathbb{Q}(\alpha))$ is trivial.

5.122. Take any $k, n \in \mathbb{N}$ with $n \geq 3$ and $gcd(k, n) = 1$.

(i) Prove that

$$\big[\mathbb{Q}_n : \mathbb{Q}\big(cos(\tfrac{2\pi k}{n})\big)\big] \;=\; 2, \quad \mathbb{Q}\big(cos(\tfrac{2\pi k}{n})\big) = \mathbb{Q}_n \cap \mathbb{R},$$
$$\text{and} \quad \big[\mathbb{Q}\big(cos(\tfrac{2\pi k}{n})\big) : \mathbb{Q}\big] \;=\; \varphi(n)/2.$$

(ii) Determine $\big[\mathbb{Q}\big(sin(\tfrac{2\pi k}{n})\big) : \mathbb{Q}\big]$.

5.123. For an integer $n \geq 3$, recall from §5.2 that a regular n-gon is constructible (by compass and straightedge) iff its central angle $2\pi/n$ is a constructible angle, iff $cos(\frac{2\pi}{n})$ is a constructible number. Prove that

$$\text{a regular } n\text{-gon is constructible} \;\; \text{iff} \;\; \varphi(n) \text{ is a power of 2.} \tag{5.35}$$

(Hint: Use part (i) of the preceding problem.)

Note: The formula (1.11) for $\varphi(n)$ in terms of the prime factorization of n shows that $\varphi(n)$ is a 2-power iff $n = 2^r p_1 \dots p_k$ for some integer $r \geq 0$ and distinct odd primes p_i such that each $\varphi(p_i)$ is a 2-power. An odd prime p with $\varphi(p)$ a 2-power is called a *Fermat prime*. If p is a Fermat prime, then $p = \varphi(p) + 1 = 2^s + 1$ for some $s \in \mathbb{N}$. The integer s must also be a 2-power, as $(2^a + 1) \big| (2^{ab} + 1)$ for $a, b \in \mathbb{N}$ with b odd. The only known Fermat primes are $3 = 2^{2^0} + 1$, $5 = 2^{2^1} + 1$, $17 = 2^{2^2} + 1$, $257 = 2^{2^3} + 1$, and $65,537 = 2^{2^4} + 1$.

Discriminant of a polynomial. Let f be a separable polynomial in $F[X]$, let $n = deg(f)$, and let E be a splitting field of f over F; so E is Galois over F. The *Galois group of f over F* is defined to be

$$\mathcal{G}(f; F) \;=\; \mathcal{G}(E/F) \;\; \text{where } E \text{ is a splitting field of } f \text{ over } F. \tag{5.36}$$

Note that $\mathcal{G}(f; F)$ is (up to isomorphism) independent of the choice of E, since E is unique up to F-isomorphism; see Note 5.53. Let $\alpha_1, \dots, \alpha_n$ be the distinct roots of f in E. Since any $\tau \in \mathcal{G}(f; F)$ permutes the α_i there is a well-defined map

$$\theta\colon \mathcal{G}(f; F) \to S_n \;\; \text{given by} \;\; \tau \mapsto \sigma \text{ where } \tau(\alpha_i) = \alpha_{\sigma(i)} \text{ for all } i. \tag{5.37}$$

Clearly, θ is a group homomorphism; moreover, θ is injective as $E = F(\alpha_1, \ldots, \alpha_n)$.

Now assume further that $char(F) \neq 2$, and that $n \geq 2$, and let

$$\Delta \;=\; \prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i) \;\in\; E^*.$$

Note that for any $\sigma \in S_n$,

$$\prod_{1 \leq i < j \leq n} (\alpha_{\sigma(j)} - \alpha_{\sigma(i)}) \;=\; sgn(\sigma)\, \Delta. \tag{5.38}$$

In particular, for any $\tau \in \mathcal{G}(f; F)$, for the θ of (5.37),

$$\tau(\Delta) \;=\; sgn(\theta(\tau))\, \Delta. \tag{5.39}$$

The *discriminant of* f is defined to be

$$disc(f) \;=\; \Delta^2. \tag{5.40}$$

Formula (5.39) shows that $disc(f) \in \mathcal{F}(\mathcal{G}(f; F)) = F$. It shows further that

$$im(\theta) \;\subseteq\; A_n \quad \text{iff} \quad \Delta \in \mathcal{F}(\mathcal{G}(f; F)) = F, \quad \text{iff} \quad disc(f) \in F^2.$$

In fact,

$$F\big(\sqrt{disc(f)}\big) \;=\; F(\Delta) \;=\; \mathcal{F}(\theta^{-1}(A_n)). \tag{5.41}$$

Example 5.124. Since $disc(f)$ is a *symmetric* polynomial of the roots of f (i.e., invariant under all permutations of the roots) it is actually expressible as a polynomial in the coefficients of f. For example, one can calculate that

$$\text{if } f \;=\; X^2 + bX + c, \quad \text{then} \quad disc(f) \;=\; b^2 - 4c,$$

and

$$\text{if } f \;=\; X^3 + aX^2 + bX + c,$$
$$\text{then} \quad disc(f) \;=\; a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc.$$

Note that if f is irreducible of degree 3, then $\mathcal{G}(f; F) \cong S_3$ or $\cong A_3$; which possibility occurs depends on whether $disc(f) \in F^2$. Also, when $\mathcal{G}(f; F) \cong S_3$, the unique quadratic extension of F in a splitting field of f over F is $F(\sqrt{disc(f)})$.

5.125. Let F be a subfield of $\mathbb{R}$ and let f be irreducible of degree 3 in $F[X]$. We know that either f splits over $\mathbb{R}$ or it has one root in $\mathbb{R}$ and two complex conjugate nonreal roots in $\mathbb{C}$. Prove that f splits over $\mathbb{R}$ iff $disc(f) > 0$.

5.126. Let f be monic and separable in $F[X]$ with $deg(f) = n \geq 2$. Let $\alpha_1, \ldots, \alpha_n$ be all the distinct roots of f in a field $K \supseteq F$ over which f splits. Let f' be the formal derivative of f. Prove that

$$disc(f) \;=\; (-1)^{n(n-1)/2} \prod_{i=1}^{n} f'(\alpha_i).$$

5.127. Let p be any odd prime number.

(i) Let $f = X^p - 1 \in \mathbb{Q}[X]$. Compute $disc(f)$ using the preceding problem.

(ii) Let $\mathbb{Q}_p$ be the p-th cyclotomic extension of $\mathbb{Q}$, which is a splitting field of $X^p - 1$ over $\mathbb{Q}$. We know (see (5.33)) that $\mathcal{G}(\mathbb{Q}_p/\mathbb{Q}) \cong \mathbb{Z}_p^*$, which is a cyclic group of order $p-1$. Prove that there is a unique field L with $\mathbb{Q} \subseteq L \subseteq \mathbb{Q}_p$ and $[L{:}\mathbb{Q}] = 2$.

(iii) For the L of part (ii), prove that

$$L \;=\; \mathbb{Q}\big(\sqrt{(-1)^{(p-1)/2}\, p}\,\big).$$

That is, $L = \mathbb{Q}(\sqrt{p}\,)$ if $p \equiv 1 \ (mod\ 4)$, while $L = \mathbb{Q}(\sqrt{-p}\,)$ if $p \equiv 3 \ (mod\ 4)$.

Note: For p an odd prime and any $n \in \mathbb{N}$, the group $\mathbb{Z}_{p^n}^*$ is cyclic (see problem 2.27(i)). Hence, there is a unique quadratic extension field L of $\mathbb{Q}$ lying in $\mathbb{Q}_{p^n}$. This is the same field as the L for $\mathbb{Q}_p$ in part (iii) above, as $\mathbb{Q}_p \subseteq \mathbb{Q}_{p^n}$. For the prime 2, we have $\mathbb{Q}_4 = \mathbb{Q}(\sqrt{-1}\,)$. But for $n \geq 3$, the noncyclic group $\mathbb{Z}_{2^n}^*$ (see problem 2.27(ii)) has three subgroups of index 2. Correspondingly, there are three subfields L of $\mathbb{Q}_{2^n}$ with $[L{:}\mathbb{Q}] = 2$, namely $\mathbb{Q}(\sqrt{-1}\,)$, $\mathbb{Q}(\sqrt{2}\,)$, and $\mathbb{Q}(\sqrt{-2}\,)$.

Quadratic reciprocity. Let p be an odd prime number. An integer n relatively prime to p is said to be a *quadratic residue mod p* if there is an integer solution x to the congruence

$$x^2 \;\equiv\; n \ (mod\ p),$$

i.e., the image $[n]_p$ of n in $\mathbb{Z}_p^*$ lies in $(\mathbb{Z}_p^*)^2$. Since the multiplicative group $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{[0]_p\}$ of the field $\mathbb{Z}_p$ is a cyclic group of even order $p-1$, its subgroup $(\mathbb{Z}_p^*)^2$ is its unique subgroup of index 2, and

$$(\mathbb{Z}_p^*)^2 \;=\; \big\{[a]_p \in \mathbb{Z}_p^* \;\big|\; [a]_p^{(p-1)/2} = [1]_p\big\}. \tag{5.42}$$

To facilitate analyzing quadratic residues, one defines the *Legendre symbol* $\left(\frac{n}{p}\right)$ by (for $n \in \mathbb{Z}$ with $p \nmid n$)

$$\left(\tfrac{n}{p}\right) = \begin{cases} 1, & \text{if } n \text{ is a quadratic residue } mod\ p, \text{ i.e., } [n]_p \in (\mathbb{Z}_p^*)^2; \\ -1, & \text{if } n \text{ is a quadratic } non\text{residue } mod\ p, \text{ i.e., } [n]_p \notin (\mathbb{Z}_p^*)^2. \end{cases} \tag{5.43}$$

Note that the Legendre symbol is multiplicative in n, i.e.,

$$\left(\tfrac{mn}{p}\right) = \left(\tfrac{m}{p}\right)\left(\tfrac{n}{p}\right) \quad \text{for } m, n \in \mathbb{Z}, \text{ each prime to } p. \tag{5.44}$$

(In particular, this says that the product of two quadratic nonresidues is a quadratic residue.) This holds because the Legendre symbol is the composition of the maps $\mathbb{Z} \setminus p\mathbb{Z} \to \mathbb{Z}_p^*$, $\mathbb{Z}_p^* \to \mathbb{Z}_p^*/(\mathbb{Z}_p^*)^2$, and $\mathbb{Z}_p^*/(\mathbb{Z}_p^*)^2 \to \{1, -1\}$, each of which is multiplicative. Because of the multiplicativity, to compute $\left(\frac{n}{p}\right)$, it suffices to compute $\left(\frac{q}{p}\right)$ for each prime q and $\left(\frac{-1}{p}\right)$. This is facilitated by the *Law of Quadratic Reciprocity*, which is one of the great gems of number theory. First discovered by Euler (but first fully proved by Gauss), the Law of Quadratic Reciprocity says that for odd primes p and q,

$$\left(\tfrac{q}{p}\right) = \begin{cases} \left(\frac{p}{q}\right), & \text{if } p \equiv 1 \ (mod\ 4) \text{ or } q \equiv 1 \ (mod\ 4); \\ -\left(\frac{p}{q}\right), & \text{if } p \equiv 3 \ (mod\ 4) \text{ and } q \equiv 3 \ (mod\ 4). \end{cases} \tag{5.45}$$

Restated,

$$\left(\tfrac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}\left(\tfrac{p}{q}\right) \text{ for any odd primes } p \text{ and } q. \tag{5.46}$$

There are many proofs of Quadratic Reciprocity. The next problem gives a proof using Galois theory and discriminants. For this, we will need

$$\left(\tfrac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} 1, & \text{if } p \equiv 1 \ (mod\ 4); \\ -1, & \text{if } p \equiv 3 \ (mod\ 4). \end{cases} \tag{5.47}$$

(Proof: Since $[-1]_p$ is the unique element of order 2 in $\mathbb{Z}_p^*$, it is a square iff the cyclic group $\mathbb{Z}_p^*$ has an element of order 4, iff $4 \,\big|\, |\mathbb{Z}_p^*| = p - 1$.)

5.128. Let p and q be odd prime numbers. Let E be a splitting field of $f = X^q - 1$ over $\mathbb{F}_p = \mathbb{Z}_p$, and let $r = [E{:}\mathbb{F}_p]$; so $E \cong \mathbb{F}_{p^r}$. Recall from problem 5.86 that r is the order of $[p]_q$ in the group $\mathbb{Z}_q^*$, and that $X^q - 1 = (X-1)g_1 \dots g_s$ in $\mathbb{F}_p[X]$, where each g_i is irreducible of degree r; so,

$$rs = q - 1.$$

(i) Prove that $[p]_q \in (\mathbb{Z}_q^*)^2$ iff the order r of $[p]_q$ in $\mathbb{Z}_q^*$ divides $(q-1)/2$, iff $2|s$. (See equation (5.42).) Restated,

$$\left(\tfrac{p}{q}\right) = (-1)^s.$$

(ii) Let $\{\alpha_1, \alpha_2, \ldots, \alpha_q\}$ be the roots of $X^q - 1$ in E, and let $\theta\colon \mathcal{G}(E/\mathbb{F}_p) \to S_q$ be the homomorphism determined by the permutation action of $\mathcal{G}(E/\mathbb{F}_p)$ on the α_i, as in (5.37). Let τ be a generator of the cyclic group $\mathcal{G}(E/\mathbb{F}_p)$, and recall from problem 5.86(iv) that $\theta(\tau)$ is a product of s disjoint cycles each of length r. Deduce that $\theta(\mathcal{G}(E/F_p)) \subseteq A_q$ (the alternating group of degree q) iff $\theta(\tau)$ is an even permutation, iff s is even. (Note that r and s cannot both be odd, as q is odd.)

(iii) Compute $disc(f)$. (This is analogous to the calculation in problem 5.127(i).) Recall that $disc(f) \in (\mathbb{F}_p^*)^2$ iff $im(\theta) \subseteq A_q$. Deduce that

$$\left(\tfrac{(-1)^{(q-1)/2}q}{p}\right) = (-1)^s.$$

(iv) By combining the results of parts (i) and (iii) and formula (5.47), deduce that

$$\left(\tfrac{q}{p}\right)\left(\tfrac{p}{q}\right) = (-1)^{(p-1)(q-1)/4},$$

which is the Law of Quadratic Reciprocity.

Note: To complete the picture on Legendre symbols, we need a formula for $\left(\frac{2}{p}\right)$ for an odd prime p. In fact,

$$\left(\tfrac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} 1, & \text{if } p \equiv 1 \text{ or } p \equiv 7 \ (mod\ 8); \\ -1, & \text{if } p \equiv 3 \text{ or } p \equiv 5 \ (mod\ 8). \end{cases} \tag{5.48}$$

Proof: Let $K = \mathbb{F}_p(\omega)$, where ω is a primitive 8-th root of unity, and let $\beta = \omega + \omega^{-1} \in K$. Since $\omega^4 = -1$, we have

$$\beta^2 = \omega^2 + 2 + \omega^{-2} = (\omega^4 + 1)\omega^{-2} + 2 = 2.$$

So, the square roots of 2 in K are β and $-\beta$. Thus,

$$\left(\tfrac{2}{p}\right) = 1 \ \text{ iff } \ \beta \in \mathbb{F}_p.$$

This holds iff $\beta^p = \beta$, since $\mathbb{F}_p = \{\alpha \in K \mid \alpha^p = \alpha\}$. We have

$$\beta^p = (\omega + \omega^{-1})^p = \omega^p + \omega^{-p},$$

as $char(K) = p$. If $p \equiv 1 \ (mod\ 8)$, then $\omega^p = \omega$, so $\beta^p = \beta$, hence $\left(\frac{2}{p}\right) = 1$. If $p \equiv 7 \ (mod\ 8)$, then $\omega^p = \omega^{-1}$, so $\beta^p = \omega^{-1} + \omega = \beta$, hence again $\left(\frac{2}{p}\right) = 1$. However, if $p \equiv 3 \ (mod\ 8)$, then $\omega^p = \omega^3$, so

$$\beta^p \ = \ \omega^3 + \omega^{-3} \ = \ \omega^4(\omega^{-1} + \omega^{-7}) \ = \ -\beta \neq \beta;$$

hence $\beta \notin \mathbb{F}_p$, so $\left(\frac{2}{p}\right) = -1$. Likewise, if $p \equiv 5 \ (mod\ 8)$, then $\omega^p = \omega^5$, so

$$\beta^p \ = \ \omega^5 + \omega^{-5} \ = \ \omega^4(\omega + \omega^{-9}) \ = \ -\beta \ \neq \beta;$$

hence, $\beta \notin \mathbb{F}_p$ and $\left(\frac{2}{p}\right) = -1$.

5.10. Radical extensions, norms, and traces

Norm and trace. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$. For any $\alpha \in K$ there as an associated F-linear transformation $L_{K,\alpha} \in \mathcal{L}_F(K, K)$ of multiplication by α, i.e, $L_{K,\alpha} \colon K \to K$ is given by

$$L_{K,\alpha}(\beta) \ = \ \alpha\beta.$$

The *norm from K to F* and *trace from K to F* of α are defined respectively by

$$N_{K/F}(\alpha) \ = \ det(L_{K,\alpha}) \quad \text{and} \quad tr_{K/F}(\alpha) \ = \ tr(L_{K,\alpha}). \tag{5.49}$$

Thus, $N_{K/F}$ and $tr_{K/F}$ are functions from K to F. Moreover, for $\alpha, \beta \in K$ and $c \in F$, since $L_{K,\alpha\beta} = L_{K,\alpha} \circ L_{K,\beta}$ and $L_{K,c} = c\, id_K$,we have

$$N_{K/F}(\alpha\beta) \ = \ N_{K/F}(\alpha)\, N_{K/F}(\beta) \ \text{ and } \ N_{K/F}(c\,\alpha) \ = \ c^n\, N_{K/F}(\alpha).$$

Indeed, since further $N_{K/F}(1) = det(id_K) = 1$, $N_{K/F}$ is a multiplicative group homomorphism from K^* to F^*. Similarly, for any $c, d \in F$ and $\alpha, \beta \in K$ since $L_{K,c\alpha+d\beta} = c\, L_{K,\alpha} + d\, L_{K,\beta}$, we have

$$tr_{K/F}(c\alpha + d\beta) \ = \ c\, tr_{K/F}(\alpha) + d\, tr_{K/F}(\beta).$$

5.129. For $F \subseteq K$ fields with $[K{:}F] < \infty$, fix $\alpha \in K$. Since the multiplication-by-α map $L_{K,\alpha} \colon K \to K$ is F-linear, we can consider F-vector space direct sum decompostions of K into $L_{K,\alpha}$-cyclic subspaces, as in §4.9.

(i) For any $\gamma \in K^*$, prove that the $L_{K,\alpha}$-cyclic subspace of K generated by γ is

$$Z(L_{K,\alpha};\gamma) = F(\alpha)\gamma = \{\beta\gamma \mid \beta \in F(\alpha)\},$$

and that the $L_{K,\alpha}$-annihilator $m_{L_{K,\alpha},\gamma}$ of γ equals the minimal polynomial $m_{F,\alpha}$.

(ii) Let $\{\gamma_1, \ldots, \gamma_k\}$ be an $F(\alpha)$-vector space base of K. So, $k = n/d$, where $n = [K{:}F]$ and $d = [F(\alpha){:}F] = deg(m_{F,\alpha})$. Prove that K has the $L_{K,\alpha}$-cyclic direct sum decomposition

$$K = Z(L_{K,\alpha};\gamma_1) \oplus \ldots \oplus Z(L_{K,\alpha};\gamma_i) \oplus \ldots \oplus Z(L_{K,\alpha};\gamma_k).$$

Deduce that there are k invariant factors of $L_{K,\alpha}$, each of which is $m_{F,\alpha}$.

(iii) Let $m_{F,\alpha} = X^d + c_{d-1}X^{d-1} + \ldots + c_iX^i + \ldots + c_0 \in F[X]$. Prove that

$$N_{K/F}(\alpha) = (-1)^n c_0^{n/d} \text{ and } tr_{K/F}(\alpha) = -(n/d)c_{d-1}. \quad (5.50)$$

(Recall (4.74).)

5.130. Let $F \subseteq K$ be fields with K Galois over F, and let $\alpha \in K$. Prove that

$$N_{K/F}(\alpha) = \prod_{\tau \in \mathcal{G}(K/F)} \tau(\alpha) \quad \text{and} \quad tr_{K/F}(\alpha) = \sum_{\tau \in \mathcal{G}(K/F)} \tau(\alpha). \quad (5.51)$$

(Recall problem 5.84(i).)

5.131. Take any $a \in F$.

(i) Let p be an odd prime number. Prove that if $X^p - a$ is irreducible in $F[X]$, then $X^{p^n} - a$ is irreducible in $F[X]$ for every $n \in \mathbb{N}$. (Hint: For $n > 1$, let β be a root of $X^{p^{n-1}} - a$. If $X^p - \beta$ is reducible in $F(\beta)[X]$, then there is $\alpha \in F(\beta)$ with $\alpha^p = \beta$. Consider $N_{F(\beta)/F}(\alpha)$.)

(ii) Prove that if $X^4 - a$ is irreducible in $F[X]$ then $X^{2^n} - a$ is irreducible in $F[X]$ for every integer $n \geq 2$. (Hint: Argue as in part (i), but consider $N_{F(\beta)/F(\beta^{2^{n-2}})}(\alpha)$.)

Note 5.132. By problem 5.51 if $char(F) \neq 2$, then $X^4 - a$ is irreducible in $F[X]$ iff $a \notin F^2$ (i.e., $X^2 - a$ is irreducible in $F[X]$) *and* $-4a \notin F^4$. Thus, by combining this with problems 5.45, 5.47, 5.48, and 5.131,

we can conclude that for any field F, any $a \in F$, and any $n \in \mathbb{N}$, the polynomial $X^n - a$ is irreducible in $F[X]$ iff for each prime p dividing n we have $a \notin F^p$ and, if $4|n$, $-4a \notin F^4$.

5.133. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$, and let E be a field containing K with E normal over F. Let $\tau_1, \ldots, \tau_k$ be all the distinct F-homomorphisms of K to E. Let $G = \mathcal{G}(E/F)$ and $H = \mathcal{G}(E/K) \subseteq G$.

(i) Prove that the τ_i are in one-to-one correspondence with the left cosets of H in G. Hence, $k = |G{:}H|$.

(ii) Prove that $k \,\big|\, [K{:}F]$.

(iii) Prove that for any $\alpha \in K$,

$$N_{K/F}(\alpha) = \Big[\prod_{i=1}^{k} \tau_i(\alpha)\Big]^{[K:F]/k} \text{ and } tr_{K/F}(\alpha) = \tfrac{[K:F]}{k} \sum_{i=1}^{k} \tau_i(\alpha) \tag{5.52}$$

5.134. *Transitivity of the norm and trace.* Let $F \subseteq L \subseteq K$ be fields with $[K{:}F] < \infty$. Prove the "transitivity formulas" for norm and trace: For any $\alpha \in K$,

$$N_{K/F}(\alpha) = N_{L/F}(N_{K/L}(\alpha)) \text{ and } tr_{K/F}(\alpha) = tr_{L/F}(tr_{K/L}(\alpha)). \tag{5.53}$$

5.135. Suppose that $char(F) \neq 2$, and let $K = F(\sqrt{a}, \sqrt{b}\,)$ for some $a, b \in F$, with $[K{:}F] = 4$. Prove that

$$F^2 \cdot im(N_{K/F}) = im(N_{F(\sqrt{a}\,)/F}) \cap im(N_{F(\sqrt{b}\,)/F}). \tag{5.54}$$

(Hint: For the inclusion $\supseteq$, take any $\gamma \in F(\sqrt{a}\,)$ such that $N_{F(\sqrt{a}\,)/F}(\gamma) \in im(N_{F(\sqrt{b}\,)/F})$, and show there is $\delta \in K^*$ with $\gamma N_{K/F(\sqrt{b}\,)}(\delta) \in F$.)

Linear independence of automorphisms. Let K be a field, and let $\tau_1, \tau_2, \ldots, \tau_n$ be distinct (ring) automophisms of K. Then, $\tau_1, \ldots, \tau_n$ are K-linearly independent in $Hom(K, K)$. That is, for any $c_1, \ldots, c_n$ in K, if $\sum_{i=1}^{n} c_i \tau_i(\alpha) = 0$ for all $\alpha \in K$, then $c_1 = \ldots = c_n = 0$.

Proof: If not, choose k minimal such that $\tau_1, \ldots, \tau_k$ are K-dependent. Thus, there are $c_1, \ldots, c_k \in K$ such that $\sum_{i=1}^{k} c_i \tau_i = 0$ and not all $c_i = 0$. Then: $k > 1$ as $\tau_1 \neq 0$; $c_k \neq 0$ by the minimality

of k; and $c_j \neq 0$ for some $j < k$ as $\tau_k \neq 0$. Choose $\beta \in K$ with $\tau_j(\beta) \neq \tau_k(\beta)$. Then, for any $\alpha \in K$, as $\tau_i(\beta\alpha) = \tau_i(\beta)\tau_i(\alpha)$,

$$\begin{aligned} 0 \;&=\; \big(\sum_{i=1}^{k} c_i\tau_i(\beta\alpha)\big) \;-\; \tau_k(\beta)\sum_{i=1}^{k} c_i\tau_i(\alpha) \\ &=\; \sum_{i=1}^{k-1}[\tau_i(\beta)-\tau_k(\beta)]c_i\tau_i(\alpha). \end{aligned}$$

Since $[\tau_j(\beta)-\tau_k(\beta)]c_j \neq 0$ and $\sum_{i=1}^{k-1}[\tau_i(\beta)-\tau_k(\beta)]c_i\tau_i = 0$, we have a contradiction to the minimality of k.

Recall *Hilbert's Theorem* 90: Let field K be a Galois extension of F with cyclic Galois group $\mathcal{G}(K/F) = \langle\sigma\rangle$. Take any $\alpha \in K^*$ with $N_{K/F}(\alpha) = 1$. Then, there is $\gamma \in K$ such that $\alpha = \sigma(\gamma)/\gamma$.

Proof: Let $n = |\,\mathcal{G}(K/F)| = |\sigma|$, and let

$$c_i \;=\; \prod_{j=0}^{i-1} \sigma^j(\alpha) \;\in\; K^* \quad \text{for} \;\; i = 1, 2, \ldots, n+1.$$

Note that $c_{i+1} = \alpha\sigma(c_i)$ and $c_n = N_{K/F}(\alpha) = 1$ by (5.51). Hence, $c_{n+1} = \alpha = c_1$. It was proved just above that the automorphisms $\sigma, \sigma^2, \ldots, \sigma^n$ are K-linearly independent. Therefore, there is $\beta \in K^*$ with $\sum_{i=1}^{n} c_i\sigma^i(\beta) \neq 0$. Let $\delta = \sum_{i=1}^{n} c_i\sigma^i(\beta) \neq 0$. Then,

$$\alpha\,\sigma(\delta) \;=\; \sum_{i=1}^{n} \alpha\sigma(c_i)\sigma^{i+1}(\beta) \;=\; \sum_{i=1}^{n} c_{i+1}\sigma^{i+1}(\beta) \;=\; \delta.$$

Hence, $\alpha = \delta/\sigma(\delta) = \sigma(\delta^{-1})/\delta^{-1}$.

5.136. Let $F \subseteq K$ be finite fields. Prove that the norm map $N_{K/F}\colon F \to K$ is surjective.

5.137. Fix an integer $n > 1$, and let F be a field containing a primitive n-th root of unity ω (so $char(F) \nmid n$).

(i) Let $K \supseteq F$ be a field with $[K{:}F] = n$, and suppose that K is Galois over F with cyclic Galois group, say $\mathcal{G}(K/F) = \langle\sigma\rangle$. Since $N_{K/F}(\omega) = \omega^n = 1$, by Hilbert 90 there is $\gamma \in K^*$ with $\omega = \sigma(\gamma)/\gamma$. Let $c = \gamma^n$. Prove that $c \in F^*$, $K = F(\gamma)$, $X^n - c$ is irreducible in $F[X]$, and K is a splitting field of $X^n - c$ over F.

(ii) Conversely, take $a \in F^*$ and let L be a splitting field of $X^n - a$ over F. Let $d = [L{:}F]$. Prove that L is Galois

over F and every irreducible factor of $X^n - a$ in $F[X]$ has degree d; hence $d|n$. Prove further that there is $b \in F$ with $b^{n/d} = a$, and that L is a splitting field over F of $X^d - b$, which is irreducible in $F[X]$. Prove also that $\mathcal{G}(L/F)$ is a cyclic group of order d (cf. problem 5.119).

5.138. Let F be a field with $char(F) = p \neq 0$. Take $c \in F$, and suppose there is no $d \in F$ with $d^p - d = c$. Let $f = X^p - X - c \in F[X]$, and let $K = F(\gamma)$ where γ is a root of f.

(i) Prove that $f = m_{F,\gamma}$, and that K is a splitting field for f over F, and that there is $\tau \in \mathcal{G}(K/F)$ with $\tau(\gamma) = \gamma + 1$. (Recall Example 5.44(iii).)

(ii) Prove that K is Galois over F and that $\mathcal{G}(K/F) = \langle\tau\rangle$, a cyclic group of order p.

Note: There is a converse to the preceding problem: Suppose that $char(F) = p \neq 0$, and let field K be a Galois extension of F with $\mathcal{G}(K/F) \cong C_p$. Then, K is a splitting field over F of $X^p - X - c$ for some $c \in F^*$. This is provable using an additive version of Hilbert's Theorem 90, which can be found, e.g., in Hungerford [**9**, Th. 7.6(i), p. 292] or Dummit & Foote [**5**, Ex. 26, p. 584].

5.139. *Kummer extensions.* Fix an integer $n > 1$, and suppose that F contains a primitive n-th root of unity ω (so $char(F) \nmid n$). Let $K \supseteq F$ be a field with $[K{:}F] < \infty$. Prove that the following conditions are equivalent:

(a) There are $c_1, \ldots, c_k \in F$ such that K is a splitting field over F of $(X^n - c_1) \ldots (X^n - c_k)$.

(b) K is Galois over F with $\mathcal{G}(K/F)$ an abelian group that is n-torsion (i.e., $\tau^n = id_F$ for every $\tau \in \mathcal{G}(K/F)$.)

A field extension K of F satisfying these equivalent conditions is called an *n-Kummer extension of F*.

5.140. *Classification of n-Kummer extensions.* Let F be a field containing a primitive n-th root of unity ω for some $n > 1$. Let K be an n-Kummer extension of F as described in the preceding problem.

Let

$$B = \{\beta \in K^* \mid \beta^n \in F\} \quad \text{and} \quad C = \{\beta^n \mid \beta \in B\} \subseteq F^*.$$

Clearly, B is a subgroup of K^* containing F^* and C is a subgroup of F^* containing F^{*n}. By the preceding problem, we have

$$K = F(\{\beta \mid \beta \in B\}).$$

(i) Prove that $B/F^* \cong C/F^{*n}$ via the map $\beta F^* \mapsto \beta^n F^{*n}$.

(ii) Let $U = \langle\omega\rangle$, a cyclic subgroup of F^* of order n. Prove that there is a well-defined bimultiplicative map

$$\psi\colon B/F^* \times \mathcal{G}(K/F) \longrightarrow U \text{ given by } \psi(\beta F^*, \tau) = \tau(\beta)/\beta.$$

(That ψ is bimultiplicative means that

$$\psi(\beta F^*, \sigma\tau) = \psi(\beta F^*, \sigma)\cdot\psi(\beta F^*, \tau)$$

and

$$\psi(\beta\gamma F^*, \tau) = \psi(\beta F^*, \tau)\cdot\psi(\gamma F^*, \tau),$$

for all $\beta, \gamma \in B$ and $\sigma, \tau \in \mathcal{G}(K/F)$.)

(iii) For any abelian n-torsion group D, let

$$D^\times = \mathit{Hom}(D, U),$$

the the abelian group of homomorphisms from D to U (as in (2.70)). Note that since D is n-torsion, $D^\times$ is isomorphic to the dual group $\mathit{Hom}(D, \Omega)$ of D as in (2.72). (For, every homomorphism from D to Ω has image in the cyclic subgroup C_n of Ω, and $C_n \cong U$.) It follows from problem 2.114(v) that

$$\text{if } |D| < \infty \text{ then } D^\times \cong D. \tag{5.55}$$

Note that the homomorphism ψ of part (ii) induces group homomorphisms

$$\mu\colon B/F^* \to (\mathcal{G}(K/F))^\times \text{ given by } \beta F^* \mapsto (\,\tau \mapsto \tau(\beta)/\beta\,)$$

and

$$\nu\colon \mathcal{G}(K/F) \to (B/F^*)^\times \text{ given by } \tau \mapsto (\,\beta F^* \mapsto \tau(\beta)/\beta\,).$$

That is,

$$\mu(\beta F^*)(\tau) = \psi(\beta F^*, \tau) = \nu(\tau)(\beta F^*).$$

Prove that μ and ν are injective. Deduce from (5.55) that $|B/F^*| < \infty$ and that μ and ν are isomorphisms. Hence,

$$B/F^* \cong (\mathcal{G}(K/F))^\times \cong \mathcal{G}(K/F)$$

and

$$|C/F^{*n}| = |B/F^*| = |\mathcal{G}(K/F)| = [K{:}F].$$

(iv) Prove that there is a one-to-one inclusion-preserving correspondence between the subgroups of B/F^* and the fields L with $F \subseteq L \subseteq K$. The correspondence is given by: For any subgroup B_0 of B containing F^*,

$$B_0/F^* \quad \longleftrightarrow \quad F(\{\beta \mid \beta \in B_0\}).$$

(v) Prove that there is a one-to-one inclusion-preserving correspondence between the finite subgroups of F^*/F^{*n} and the n-Kummer extensions of F in an algebraic closure A of F. In this correspondence, for any finite subgroup C/F^{*n} of F^*/F^{*n} (where C is a group with $F^{*n} \subseteq C \subseteq F^*$), the corresponding field is $F(\{\sqrt[n]{c} \mid c \in C\})$.

5.141. Assume that $char(F) \neq 2$. Let $L \supseteq F$ be a field with $[L{:}F] = 2$. So, $L = F(\sqrt{c}\,)$ for some $c \in F \backslash F^2$, and $\mathcal{G}(L/F) = \{id_L, \sigma\}$, where $\sigma(\sqrt{c}\,) = -\sqrt{c}\,$.

(i) Prove that the following two conditions are equivalent:
  (a) $-1 \in im(N_{L/F})$.
  (b) $c = a^2 + b^2$ for some $a, b \in F$.

(ii) Prove that there is a field $E \supseteq L$ with $[E{:}F] = 4$, and E is Galois over F with $\mathcal{G}(E/F)$ cyclic iff $-1 \in im(N_{L/F})$. (Hint: If $-1 = N_{L/F}(\gamma)$, then $\gamma^2 = \sigma(\beta)/\beta$ for some $\beta \in L^*$. Let $E = L(\sqrt{\beta}\,)$. Alternatively, apply problem 5.63(iv).)

5.142. This is a generalization of the preceding problem. Let p be a prime number, and let F be a field containing a primitive p-th root of unity ω; so, $char(F) \neq p$. Let L be a Galois extension of F with $[L{:}F] = p^n$ for some $n \in \mathbb{N}$ and $\mathcal{G}(L/F)$ cyclic. Prove that the following conditions are equivalent:

(a) $\omega \in im(N_{L/F})$.

(b) There is a field $E \supseteq L$ with $[E{:}F] = p^{n+1}$ and E Galois over F with $\mathcal{G}(E/F)$ cyclic.

5.143. *Pythagorean triples.* A triple (a, b, c) of positive integers is called a *Pythagorean triple* if $c^2 = a^2 + b^2$. (By the Pythagorean Theorem, (a, b, c) is a Pythagorean triple just when a, b, c are the integer side lengths of a right triangle, with c the hypotenuse.) Note that in such a triple, a and b cannot both be odd, since $c^2 \not\equiv 2 \ (mod\ 4)$. For example, $(3, 4, 5)$, $(5, 12, 13)$, $(63, 16, 65)$, $(33, 56, 65)$ are Pythagorean triples. Prove that (after interchanging a and b if necessary) every Pythagorean triple has the form

$$a = r(m^2 - n^2), \quad b = r(2mn), \quad c = r(m^2 + n^2), \tag{5.56}$$

for some $m, n, r \in \mathbb{N}$. (Clearly, every (a, b, c) as in (5.56) is a Pythagorean triple.) (Hint: First use Hilbert 90 to show that (a, b, c) has the form given in (5.56) for some $m, n \in \mathbb{N}$ and $r \in \mathbb{Q}$.) Note that the r cannot be omitted from the equations in (5.56), as illustrated by the Pythagorean triple $(9, 12, 15)$.

5.144. Let $F \subseteq L \subseteq K$ be fields such that K Galois is over F with $\mathcal{G}(K/F)$ a cyclic group; let $n = [K{:}L]$. For $c \in F$, prove that if c^n is a norm from K to F, then c is a norm from L to F. (Hint: Let $\mathcal{G}(K/F) = \langle\sigma\rangle$ and let $k = [L{:}F]$. Work with the "partial norm map" $N_k \colon K \to K$ given by $\alpha \mapsto \alpha\sigma(\alpha)\sigma^2(\alpha)\ldots\sigma^{k-1}(\alpha)$. Note that $N_k \circ N_{K/L} = N_{K/L} \circ N_k = N_{K/F}$.)

5.145. Let $F \subseteq K$ be fields with $[K{:}F] < \infty$.

(i) Prove that if K is not separable over F, then $tr_{K/F}(\alpha) = 0$ for all $\alpha \in K$.

(ii) Prove that if K is separable over F then there is $\alpha \in K$ with $tr_{K/F}(\alpha) \neq 0$. (Hint: First prove this for K Galois over F.)

(iii) Suppose that K is separable over F. Let $K^\times = \mathcal{L}_F(K, F)$, the dual space of K as an F-vector space. Define $\theta \colon K \to K^\times$ by

$$\theta(\alpha)(\beta) = tr_{K/F}(\alpha\beta) \quad \text{for all } \alpha, \beta \in K.$$

Prove that θ is an F-vector space isomorphism.

Ordered fields. A pair (F, P) is called an *ordered field* if F is a field and P is a subset of F^* such that

(i) if $a, b \in P$, then $a + b \in P$ and $ab \in P$; and

(ii) $F = P \cup \{0\} \cup -P$, a disjoint union, where $-P = \{-a \mid a \in P\}$.

We can then define a total ordering $<$ on F by:

$$a < b \text{ just when } b - a \in P, \quad \text{for any } a, b \in F.$$

Note that $<$ satisfies the familiar properties of the orderings on $\mathbb{R}$ and $\mathbb{Q}$: for any $a, b \in F$, $a < b$ or $a = 0$ or $b < a$ and only one of these holds. Also, if $a < b$ and $b < c$, then $a < c$; if $a < b$ and $c < d$, then $a + c < b + d$; if $a < b$ and $0 < c$, then $ac < bc$. (Conversely, given a relation $<$ on F with these properties, then for $P = \{a \in F \mid 0 < a\}$ we have (F, P) is an ordered field.) Note that if (F, P) is an ordered field, then $F^{*2} \subseteq P$. Also, $char(F) = 0$ since for any prime number p, the sum $\sum_{i=1}^{p} 1_F$ lies in P, so it is nonzero. A field with an ordering is called a *formally real* field.

5.146. *Real closed fields.* An ordered field (F, P) is said to be *real closed* if $P = F^{*2}$ and F has no proper field extensions of odd finite degree. For example, $(\mathbb{R}, \mathbb{R}^{*2})$ is real closed. The same argument as for $\mathbb{R}$ in problem 5.101 shows that if (F, P) is real closed, then $F(\sqrt{-1}\,)$ is algebraically closed. This problem asks you to prove that the only fields of characteristic 0 whose algebraic closure is a proper finite degree extension are real closed fields. For this, let F be a field with $char(F) = 0$, let A be an algebraic closure of F, and assume that $1 < [A{:}F] < \infty$.

(i) Prove that if f is irreducible in $F[X]$ then $deg(f) < [A{:}F]$.

(ii) Note that A is Galois over F. If p is a prime number dividing $[A{:}F]$, prove that there is a field L with $F \subseteq L \subseteq A$ and $[A{:}L] = p$. Then prove that $p = 2$. (Hint: If p is odd, get a contradiction to part (i) by using problems 5.137(i) and 5.131(i).)

(iii) For the L of part (ii), prove that (L, L^{*2}) is a real closed field and that $A = L(\sqrt{-1}\,)$. (Hint: Use problems 5.141 and 5.131(ii) and the note thereafter.)

(iv) Prove that $F = L$. Hence, (F, F^{*2}) is a real closed field and its algebraic closure is $F(\sqrt{-1}\,)$.

Note: It is known that if $char(F) = p \neq 0$, and $[A{:}F] < \infty$ where A is an algebraic closure of F, then $F = A$. The proof is similar to that in the preceding problem, using also problem 5.47 and the result that every Galois extension of F of degree p lies in a Galois extension of F of degree p^2. See Hungerford [**9**, Ex. 6(b), p. 297] for a proof of this result on degree p^2 extensions.

Example 5.147.

(i) If (F, F^2) is a real closed field with algebraic closure $A = F(\sqrt{-1})$ and A_0 is any algebraicaly closed subfield of A, then $\big(F \cap A_0, (F \cap A_0)^2\big)$ is a real closed field. (For example, if $A_{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, then, for the field $\mathbb{R} \cap A_{\mathbb{Q}}$ of "real algebraic numbers," $\big(\mathbb{R} \cap A_Q, (R \cap A_Q)^2\big)$ is real closed.)

(ii) It is known that if (F, P) is any ordered field and A is an algebraic closure of F, then there is a real closed field (R, R^2) with $F \subseteq R \subseteq A$ (so $A = R(\sqrt{-1})$) and $R^2 \cap F = P$; moreover, (R, R^2) is unique up to isomorphism. For more on ordered and real closed fields, see, e.g., Prestel's book [**18**].

5.11. Solvability by radicals

5.148. Let p be a prime number, and let $F \subseteq L$ be fields with $[L{:}F] < \infty$. Prove that the following conditions are equivalent:

(a) There is a field $E \supseteq L$ with E Galois over F. and $[E{:}F] = p^n$ for some $n \in \mathbb{N}$.

(b) There are fields $F = L_0 \subseteq L_1 \subseteq \ldots \subseteq L_i \subseteq \ldots \subseteq L_k = L$ such that each L_i is Galois over L_{i-1} with $[L_i{:}L_{i-1}] = p$.

(Hint: To find E given the L_i, apply problem 5.61.)

5.149. Let $\alpha \in \mathbb{R}$ with α algebraic over $\mathbb{Q}$, and let K be a splitting field of $m_{\mathbb{Q},\alpha}$ over $\mathbb{Q}$. Prove that α is a constructible number as in §5.2 iff $[K{:}\mathbb{Q}]$ is a power of 2. (Hint: Use the preceding problem.)

Solvability by radicals. Let f be nonconstant in $F[X]$. Then f is said to be *solvable by radicals* over F if there is a chain of fields

$F = L_0 \subseteq L_1 \subseteq \ldots \subseteq L_k$ such that for each $i \geq 1$, $L_i = L_{i-1}(\alpha_i)$ where $\alpha_i^{n_i} \in L_{i-1}$ for some $n_i \in \mathbb{N}$, and f splits over L_k.

Recall *Galois's Theorem on Sovability by Radicals*: Suppose that $char(F) = 0$ and that f is nonconstant in $F[X]$. Then, f is solvable by radicals iff its Galois group $\mathcal{G}(f;F)$ is a solvable group. See any text covering Galois theory for a proof of this.

5.150. Suppose that $char(F) = p \neq 0$, and suppose that there is $c \in F$ with $c \neq d^p - d$ for each $d \in F$. Let $f = X^p - X - c \in F[X]$, and let $K = F(\alpha)$ where α is a root of f. Recall from problem 5.138 that K is a splitting field of f over F, and that K is Galois over F with $\mathcal{G}(K/F) \cong C_p$. Prove that even though $\mathcal{G}(f;F)$ is a solvable group, f is not solvable by radicals over F. (This shows the need for assuming $char(F) = 0$ in Galois's Theorem.)

5.151. Let F be a subfield of $\mathbb{R}$, and let f be irreducible in $F[X]$ with $deg(f) = p$, a prime number. Suppose that f has exactly $p-2$ distinct roots in $\mathbb{R}$ (so f has a single pair of complex conjugate roots in $\mathbb{C} \setminus \mathbb{R}$). Prove that $\mathcal{G}(f;F) \cong S_p$. Thus, f is not solvable by radicals if $p \geq 5$. (Hint: Prove that the image of $\mathcal{G}(f;F)$ in S_p contains a p-cycle and a transposition; then use problem 2.41.)

5.152. Let p be an odd prime number. Let

$$f \;=\; 1 + X(X^2 + p - 1) \prod_{i=2}^{p-2} (X - i) \in \mathbb{Z}[X].$$

Prove that f is irreducible in $\mathbb{Q}[X]$ and that $\mathcal{G}(f;\mathbb{Q}) \cong S_p$. So, f is not solvable by radicals over $\mathbb{Q}$ if $p \geq 5$.

5.153. Suppose that $char(F) = 0$. Take any irreducible f in $F[X]$, such that $deg(f)$ is a prime number p. Let E be a splitting field of f over F. Prove the theorem of Galois that f is solvable by radicals iff $E = F(\alpha, \beta)$ for any two distinct roots of f in E. (Hint: Use problem 2.98.)

5.154. Let F be a subfield of $\mathbb{R}$, and let $f = X^3 + bX + c$ be irreducible in $F[X]$. Recall that $disc(f) = -4b^3 - 27c^2$ (see Example 5.124). Cardan's formula (see, e.g., Cox [**4**, pp. 4–5] or Tignol [**23**, pp. 15–16]) says that the roots α of f are given by $\alpha = \beta - b/(3\beta)$, where

$$\beta \;=\; \sqrt[3]{\tfrac{1}{2}\big(-c \pm \sqrt{c^2 + 4b^3/27}\,\big)} \;=\; \sqrt[3]{\tfrac{1}{2}\big(-c \pm \sqrt{-disc(f)/27}\,\big)}\,.$$

Thus, when f has all three roots in $\mathbb{R}$ (so $disc(f) > 0$ by problem 5.125) the formula for its roots involves the nonreal number $\sqrt{-disc(f)}$. This is actually unavoidable: Prove that if f has three roots in $\mathbb{R}$, then f is not "solvable by real radicals," i.e., there is no chain of fields $F = L_0 \subseteq L_1 \subseteq \ldots \subseteq L_k$ with $L_k \subseteq \mathbb{R}$ such that for each $i \geq 1$, $L_i = L_{i-1}(\alpha_i)$ where $\alpha_i^{n_i} \in L_{i-1}$ for some $n_i \in \mathbb{N}$, and f has a root in L_k.

Suggestions for Further Reading

Here are some suggestions for collateral reading or deeper study in various areas of abstract algebra.

There are a number of very good texts in abstract algebra. These include Artin [**1**], Dummit & Foote [**5**], Hungerford [**9**], Jacobson [**10**], Knapp [**13**], and Lang [**16**]. Dummit & Foote and Hungerford have particularly extensive problem sets. For more on group theory, see Rotman [**20**] or Hall [**7**].

In ring and module theory, for commutative rings see Atiyah & MacDonald [**2**], and for noncommutative rings, see Lam [**15**].

For linear algebra, see Hoffman & Kunze [**8**].

Two outstanding texts on Galois theory are the books by Cox [**4**] and by Tignol [**23**]. Each has interesting historical commentary on the work of Galois and his predecessors.

For algebraic number theory, there are many good texts, e.g., Marcus [**17**] and Weiss [**24**]. Marcus's book has an outstanding selection of problems.

Two important more advanced areas of algebra not treated here are homological algebra and algebraic geometry. See Rotman [**22**] for a good introduction to homological algebra. The book by Reid [**19**] provides a gentle introduction to algebraic geometry; see the references provided there for further reading.

Bibliography

When available, *Mathematical Reviews* reference numbers are indicated at the end of each bibliographic entry as MR*******. See www.ams.org/mathscinet.

1. Artin, Michael, *Algebra*, second ed., Prentice Hall, Boston, MA, 2011.
2. Atiyah, Michael F. and Macdonald, Ian G., *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969. MR0242802
3. Borevich, Zenon I. and Shafarevich, Igor R., *Number theory*, Pure and Applied Mathematics, Vol. 20, Academic Press, New York-London, 1966. MR0195803
4. Cox, David A., *Galois theory*, second ed., Pure and Applied Mathematics (Hoboken), John Wiley & Sons, Inc., Hoboken, NJ, 2012. MR2919975
5. Dummit, David S. and Foote, Richard M., *Abstract algebra*, third ed., John Wiley & Sons, Inc., Hoboken, NJ, 2004. MR2286236
6. Ebbinghaus, Heinz-Dieter; Hermes, Hans; Hirzebruch, Friedrich; Koecher, Max; Mainzer, Klaus; Neukirch, Jürgen; Prestel, Alexander; and Remmert, Reinhold, *Numbers*, Graduate Texts in Mathematics, vol. 123, Springer-Verlag, New York, 1990, Readings in Mathematics. MR1066206
7. Hall, Jr., Marshall, *The theory of groups*, Chelsea Publishing Co., New York, 1976, Reprint of the 1968 edition. MR0414669
8. Hoffman, Kenneth and Kunze, Ray, *Linear algebra*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. MR0276251

9. Hungerford, Thomas W., *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York-Berlin, 1980, Reprint of the 1974 original. MR600654
10. Jacobson, Nathan, *Basic algebra. I*, second ed., W. H. Freeman and Company, New York, 1985. MR780184
11. Janusz, Gerald J., *Algebraic number fields*, second ed., Graduate Studies in Mathematics, vol. 7, American Mathematical Society, Providence, RI, 1996. MR1362545
12. Kaplansky, Irving, *Set theory and metric spaces*, second ed., Chelsea Publishing Co., New York, 1977. MR0446980
13. Knapp, Anthony W., *Basic algebra*, Birkhäuser Boston, Inc., Boston, MA, 2006. MR2257570
14. Lam, Tsit Yuen and Leep, David B., *Combinatorial structure on the automorphism group of S_6*, Exposition. Math. **11** (1993), no. 4, 289–308. MR1240362
15. Lam, Tsit Yuen, *A first course in noncommutative rings*, second ed., Graduate Texts in Mathematics, vol. 131, Springer-Verlag, New York, 2001. MR1838439
16. Lang, Serge, *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR1878556
17. Marcus, Daniel A., *Number fields*, Springer-Verlag, New York-Heidelberg, 1977, Universitext. MR0457396
18. Prestel, Alexander, *Lectures on formally real fields*, Lecture Notes in Mathematics, vol. 1093, Springer-Verlag, Berlin, 1984. MR769847
19. Reid, Miles, *Undergraduate algebraic geometry*, London Mathematical Society Student Texts, vol. 12, Cambridge University Press, Cambridge, 1988. MR982494
20. Rotman, Joseph J., *An introduction to the theory of groups*, third ed., Allyn and Bacon, Inc., Boston, MA, 1984. MR745804
21. ———, *Advanced modern algebra*, Prentice Hall, Inc., Upper Saddle River, NJ, 2002. MR2043445
22. ———, *An introduction to homological algebra*, second ed., Universitext, Springer, New York, 2009. MR2455920
23. Tignol, Jean-Pierre, *Galois' theory of algebraic equations*, second ed., World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2016. MR3444922
24. Weiss, Edwin, *Algebraic number theory*, Dover Publications, Inc., Mineola, NY, 1998, Reprint of the 1963 original. MR1635455

Index of Notation

Problem numbers are given in **bold** face.

Subject and Terminology Index

Problem numbers are given in **bold** face.

Selected Published Titles in This Series

82 **A. R. Wadsworth,** Problems in Abstract Algebra, 2017

80 **Matt DeVos and Deborah A. Kent,** Game Theory, 2016

79 **Kristopher Tapp,** Matrix Groups for Undergraduates, Second Edition, 2016

78 **Gail S. Nelson,** A User-Friendly Introduction to Lebesgue Measure and Integration, 2015

77 **Wolfgang Kühnel,** Differential Geometry: Curves — Surfaces — Manifolds, Third Edition, 2015

76 **John Roe,** Winding Around, 2015

75 **Ida Kantor, Jiří Matoušek, and Robert Šámal,** Mathematics++, 2015

74 **Mohamed Elhamdadi and Sam Nelson,** Quandles, 2015

73 **Bruce M. Landman and Aaron Robertson,** Ramsey Theory on the Integers, Second Edition, 2014

72 **Mark Kot,** A First Course in the Calculus of Variations, 2014

71 **Joel Spencer,** Asymptopia, 2014

70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014

69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014

68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013

67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013

66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013

65 **Victor H. Moll,** Numbers and Functions, 2012

64 **A. B. Sossinsky,** Geometries, 2012

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012

62 **Rebecca Weber,** Computability Theory, 2012

61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011

60 **Richard Evan Schwartz,** Mostly Surfaces, 2011

59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011

For a complete list of titles in this series, visit the AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.

This is a book of problems in abstract algebra for strong undergraduates or beginning graduate students. It can be used as a supplement to a course or for self-study. The book provides more variety and more challenging problems than are found in most algebra textbooks. It is intended for students wanting to enrich their learning of mathematics by tackling problems that take some thought and effort to solve. The book contains problems on groups (including the Sylow Theorems, solvable groups, presentation of groups by generators and relations, and structure and duality for finite abelian groups); rings (including basic ideal theory and factorization in integral domains and Gauss's Theorem); linear algebra (emphasizing linear transformations, including canonical forms); and fields (including Galois theory). Hints to many problems are also included.

ISBN 978-1-4704-3583-7

9 781470 435837

STML/82

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-82