

A BOOK OF ABSTRACT ALGEBRA

Second Edition

Charles C. Pinter
Professor of Mathematics
Bucknell University

Dover Publications, Inc., Mineola, New York

Copyright © 1982, 1990 by Charles C. Pinter
All rights reserved.

Bibliographical Note

This Dover edition, first published in 2010, is an unabridged republication of the 1990 second edition of the work originally published in 1982 by the McGraw-Hill Publishing Company, Inc., New York.

Library of Congress Cataloging-in-Publication Data

Pinter, Charles C, 1932–

A book of abstract algebra / Charles C. Pinter. — Dover ed.

p. cm.

Originally published: 2nd ed. New York : McGraw-Hill, 1990.

Includes bibliographical references and index.

ISBN-13: 978-0-486-47417-5

ISBN-10: 0-486-47417-8

1. Algebra, Abstract. I. Title.

QA162.P56 2010

512'.02—dc22

2009026228

Manufactured in the United States by Courier Corporation
47417803

www.doverpublications.com

Preface

Chapter 1 Why Abstract Algebra?

History of Algebra. New Algebras. Algebraic Structures. Axioms and Axiomatic Algebra. Abstraction in Algebra.

Chapter 2 Operations

Operations on a Set. Properties of Operations.

Chapter 3 The Definition of Groups

Groups. Examples of Infinite and Finite Groups. Examples of Abelian and Nonabelian Groups. Group Tables.

Theory of Coding: Maximum-Likelihood Decoding.

Chapter 4 Elementary Properties of Groups

Uniqueness of Identity and Inverses. Properties of Inverses.

Direct Product of Groups.

Chapter 5 Subgroups

Definition of Subgroup. Generators and Defining Relations.

Cayley Diagrams. Center of a Group. Group Codes; Hamming Code.

Chapter 6 Functions

Injective, Surjective, Bijective Function. Composite and Inverse of Functions.

Finite-State Machines. Automata and Their Semigroups.

Chapter 7 Groups of Permutations

Symmetric Groups. Dihedral Groups.

An Application of Groups to Anthropology.

Chapter 8 Permutations of a Finite Set

Decomposition of Permutations into Cycles. Transpositions. Even and Odd Permutations. Alternating Groups.

Chapter 9 Isomorphism

The Concept of Isomorphism in Mathematics. Isomorphic and Nonisomorphic Groups. Cayley's Theorem.

Chapter 10 Order of Group Elements

Powers/Multiples of Group Elements. Laws of Exponents. Properties of the Order of Group Elements.

Chapter 11 Cyclic Groups

Finite and Infinite Cyclic Groups. Isomorphism of Cyclic Groups. Subgroups of Cyclic Groups.

Chapter 12 Partitions and Equivalence Relations

Chapter 13 Counting Cosets

Lagrange's Theorem and Elementary Consequences.

Survey of Groups of Order ≤ 10 .

Number of Conjugate Elements. Group Acting on a Set.

Chapter 14 Homomorphisms

Elementary Properties of Homomorphisms. Normal Subgroups. Kernel and Range.

Inner Direct Products. Conjugate Subgroups.

Chapter 15 Quotient Groups

Quotient Group Construction. Examples and Applications.

The Class Equation. Induction on the Order of a Group.

Chapter 16 The Fundamental Homomorphism Theorem

Fundamental Homomorphism Theorem and Some Consequences.

The Isomorphism Theorems. The Correspondence Theorem. Cauchy's Theorem. Sylow Subgroups. Sylow's Theorem. Decomposition Theorem for Finite Abelian Groups.

Chapter 17 Rings: Definitions and Elementary Properties

Commutative Rings. Unity. Invertibles and Zero-Divisors. Integral Domain. Field.

Chapter 18 Ideals and Homomorphisms

Chapter 19 Quotient Rings

Construction of Quotient Rings. Examples. Fundamental Homomorphism Theorem and Some Consequences. Properties of Prime and Maximal Ideals.

Chapter 20 Integral Domains

Characteristic of an Integral Domain. Properties of the Characteristic. Finite Fields. Construction of the Field of Quotients.

Chapter 21 The Integers

Ordered Integral Domains. Well-ordering. Characterization of $\mathbf{Z}$ Up to Isomorphism. Mathematical Induction. Division Algorithm.

Chapter 22 Factoring into Primes

Ideals of $\mathbb{Z}$. Properties of the GCD. Relatively Prime Integers. Primes. Euclid's Lemma. Unique Factorization.

Chapter 23 Elements of Number Theory (Optional)

Properties of Congruence. Theorems of Fermat and Euler. Solutions of Linear Congruences. Chinese Remainder Theorem.

Wilson's Theorem and Consequences. Quadratic Residues. The Legendre Symbol. Primitive Roots.

Chapter 24 Rings of Polynomials

Motivation and Definitions. Domain of Polynomials over a Field. Division Algorithm.

Polynomials in Several Variables. Fields of Polynomial Quotients.

Chapter 25 Factoring Polynomials

Ideals of $F[x]$. Properties of the GCD. Irreducible Polynomials. Unique factorization.

Euclidean Algorithm.

Chapter 26 Substitution in Polynomials

Roots and Factors. Polynomial Functions. Polynomials over $\mathbb{Q}$. Eisenstein's Irreducibility Criterion. Polynomials over the Reals. Polynomial Interpolation.

Chapter 27 Extensions of Fields

Algebraic and Transcendental Elements. The Minimum Polynomial. Basic Theorem on Field Extensions.

Chapter 28 Vector Spaces

Elementary Properties of Vector Spaces. Linear Independence. Basis. Dimension. Linear Transformations.

Chapter 29 Degrees of Field Extensions

Simple and Iterated Extensions. Degree of an Iterated Extension.

Fields of Algebraic Elements. Algebraic Numbers. Algebraic Closure.

Chapter 30 Ruler and Compass

Constructible Points and Numbers. Impossible Constructions.

Constructible Angles and Polygons.

Chapter 31 Galois Theory: Preamble

Multiple Roots. Root Field. Extension of a Field. Isomorphism.

Roots of Unity. Separable Polynomials. Normal Extensions.

Chapter 32 Galois Theory: The Heart of the Matter

Field Automorphisms. The Galois Group. The Galois Correspondence. Fundamental Theorem of Galois Theory.

Computing Galois Groups.

Chapter 33 Solving Equations by Radicals

Radical Extensions. Abelian Extensions. Solvable Groups. Insolubility of the Quintic.

Appendix A	Review of Set Theory
Appendix B	Review of the Integers
Appendix C	Review of Mathematical Induction
	Answers to Selected Exercises
	Index

* Italic headings indicate topics discussed in the exercise sections.

CHAPTER TWENTY-SEVEN

EXTENSIONS OF FIELDS

In the first 26 chapters of this book we introduced the cast and set the scene on a vast and complex stage. Now it is time for the action to begin. We will be surprised to discover that none of our effort has been wasted; for every notion which was defined with such meticulous care, every subtlety, every fine distinction will have its use and play its prescribed role in the story which is about to unfold.

We will see modern algebra reaching out and merging with other disciplines of mathematics; we will see its machinery put to use for solving a wide range of problems which, on the surface, have nothing whatever to do with modern algebra. Some of these problems—ancient problems of geometry, riddles about numbers, questions concerning the solutions of equations—reach back to the very beginnings of mathematics. Great masters of the art of mathematics puzzled over them in every age and left them unsolved, for the machinery to solve them was not there. Now, with a light touch modern algebra uncovers the answers.

Modern algebra was not built in an ivory tower but was created part and parcel with the rest of mathematics—tied to it, drawing from it, and offering it solutions. Clearly it did not develop as methodically as it has been presented here. It would be pointless, in a first course in abstract algebra, to replicate all the currents and crosscurrents, all the hits and misses and false starts. Instead, we are provided with a finished product in which the agonies and efforts that went into creating it cannot be discerned. There is a disadvantage to this: without knowing the origin of a given concept, without knowing the specific problems which gave it birth, the student often wonders what it means and why it was ever invented.

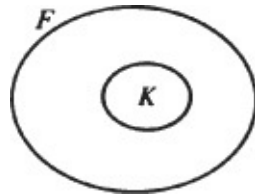
We hope, beginning now, to shed light on that kind of question, to justify what we have already done, and to demonstrate that the concepts introduced in earlier chapters are correctly designed for their intended purposes.

Most of classical mathematics is set in a framework consisting of fields, especially $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$. The theory of equations deals with polynomials over $\mathbf{R}$ and $\mathbf{C}$, calculus is concerned with functions over $\mathbf{R}$, and plane geometry is set in $\mathbf{R} \times \mathbf{R}$. It is not surprising, therefore, that modern efforts to generalize and unify these subjects should also center around the study of fields. It turns out that a great variety of problems, ranging from geometry to practical computation, can be translated into the language of fields and formulated entirely in terms of the theory of fields. The study of fields will therefore be our central

concern in the remaining chapters, though we will see other themes merging and flowing into it like the tributaries of a great river.

If F is a field, then a *subfield* of F is any nonempty subset of F which is closed with respect to addition and subtraction, multiplication and division. (It would be equivalent to say: closed with respect to addition and negatives, multiplication and multiplicative inverses.) As we already know, if K is a subfield of F , then K is a field in its own right.

If K is a subfield of F , we say also that F is an *extension field* of K . When it is clear in context that both F and K are fields, we say simply that F is an *extension* of K .



Given a field F , we may *look inward* from F at all the subfields of F . On the other hand, we may *look outward* from F at all the extensions of F . Just as there are relationships between F and its subfields, there are also interesting relationships between F and its extensions. One of these relationships, as we shall see later, is highly reminiscent of Lagrange's theorem—an inside-out version of it.

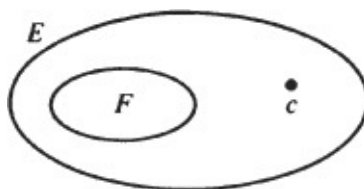
Why should we be interested in looking at the extensions of fields? There are several reasons, but one is very special. If F is an arbitrary field, there are, in general, polynomials over F which have no roots in F . For example, $x^2 + 1$ has no roots in $\mathbb{R}$. This situation is unfortunate but, it turns out, not hopeless. For, as we shall soon see, every polynomial over any field F *has roots*. If these roots are not already *in* F , they are in a suitable extension of F . For example, $x^2 + 1 = 0$ has solutions in $\mathbb{C}$.

In the matter of factoring polynomials and extracting their roots, $\mathbb{C}$ is utopia! In $\mathbb{C}$ every polynomial $a(x)$ of degree n has exactly n roots $c_1, \dots, c_n$ and can therefore be factored as $a(x) = k(x - c_1)(x - c_2) \cdots (x - c_n)$. This ideal situation is not enjoyed by all fields—far from it! In an arbitrary field F , a polynomial of degree n may have any number of roots, from no roots to n roots, and there may be irreducible polynomials of any degree whatever. This is a messy situation, which does not hold the promise of an elegant theory of solutions to polynomial equations. However, it turns out that F *always has a suitable extension* E such that any polynomial $a(x)$ of degree n over F has *exactly* n solutions in E . Therefore, $a(x)$ can be factored in $E[x]$ as

$$a(x) = k(x - c_1)(x - c_2) \cdots (x - c_n)$$

Thus, paradise is regained by the expedient of enlarging the field F . This is one of the strongest reasons for our interest in field extensions. They will give us a trim and elegant theory of solutions to polynomial equations.

Now, let us get to work! Let E be a field, F a subfield of E , and c any



element of E . We define the *substitution function* σ_c as follows:

For every polynomial $a(x)$ in $F[x]$,

$$\sigma_c(a(x)) = a(c)$$

Thus, σ_c is the function “substitute c for x .” It is a function from $F[x]$ into E . In fact, σ_c is a *homomorphism*. This is true because

$$\underbrace{\sigma_c(a(x) + b(x))}_{a(c) + b(c)} = \underbrace{\sigma_c(a(x))}_{a(c)} + \underbrace{\sigma_c(b(x))}_{b(c)}$$

and

$$\underbrace{\sigma_c(a(x)b(x))}_{a(c)b(c)} = \underbrace{\sigma_c(a(x))}_{a(c)} \underbrace{\sigma_c(b(x))}_{b(c)}$$

The kernel of the homomorphism σ_c is the set of all the polynomials $a(x)$ such that $a(c) = \sigma_c(a(x)) = 0$. That is, *the kernel of σ_c consists of all the polynomials $a(x)$ in $F[x]$ such that c is a root of $a(x)$.*

Let J_c denote the kernel of σ_c ; since the kernel of any homomorphism is an ideal, J_c is an ideal of $F[x]$.

An element c in E is called *algebraic over F* if it is the root of some nonzero polynomial $a(x)$ in $F[x]$. Otherwise, c is called *transcendental over F* . Obviously c is algebraic over F iff J_c contains nonzero polynomials, and transcendental over F iff $J_c = \{0\}$.

We will confine our attention now to the case where c is algebraic. The transcendental case will be examined in Exercise G at the end of this chapter.

Thus, let c be algebraic over F , and let J_c be the kernel of σ_c (where σ_c is the function “substitute c for x ”). Remember that in $F[x]$ every ideal is a principal ideal; hence $J_c = \langle p(x) \rangle$ = the set of all multiples of $p(x)$, for some polynomial $p(x)$. Since every polynomial in J_c is a multiple of $p(x)$, $p(x)$ is a polynomial of *lowest degree among all the nonzero polynomials in J_c* . It is easy to see that $p(x)$ is *irreducible*; otherwise we could factor it into polynomials of lower degree, say $p(x) = f(x)g(x)$. But then $0 = p(c) = f(c)g(c)$, so $f(c) = 0$ or $g(c) = 0$, and therefore either $f(x)$ or $g(x)$ is in J_c . This is impossible, because we have just seen that $p(x)$ has the *lowest degree* among all the polynomials in J_c , whereas $f(x)$ and $g(x)$ both have lower degree than $p(x)$.

Since every *constant* multiple of $p(x)$ is in J_c , we may take $p(x)$ to be monic, that is, to have leading coefficient 1. Then $p(x)$ is the *unique* monic polynomial of lowest degree in J_c . (Also, it is the only monic irreducible polynomial in J_c .) This polynomial $p(x)$ is called the *minimum polynomial* of c over F , and will be of considerable importance in our discussions in a later chapter.

Let us look at an example: $\mathbb{R}$ is an extension field of $\mathbb{Q}$, and $\mathbb{R}$ contains the irrational number $\sqrt{2}$. The function $\sigma_{\sqrt{2}}$ is the function “substitute $\sqrt{2}$ for x ”; for example $\sigma_{\sqrt{2}}(x^4 - 3x^2 + 1) = \sqrt{2}^4 - 3\sqrt{2}^2 + 1 = -1$. By our discussion above, $\sigma_{\sqrt{2}}: \mathbb{Q}[x] \rightarrow \mathbb{R}$ is a homomorphism and its kernel consists of all the polynomials in $\mathbb{Q}[x]$ which have $\sqrt{2}$ as one of their roots. The monic polynomial of least degree in $\mathbb{Q}[x]$ having $\sqrt{2}$ as a root is $p(x) = x^2 - 2$; hence $x^2 - 2$ is the minimum polynomial of $\sqrt{2}$ over $\mathbb{Q}$.

Now, let us turn our attention to the *range* of σ_c . Since σ_c is a homomorphism, its range is obviously closed with respect to addition, multiplication, and negatives, but it is not obviously closed with respect to multiplicative inverses. Not obviously, but in fact it *is* closed for multiplicative inverses, which is far from self-evident, and quite a remarkable fact. In order to prove this, let $f(c)$ be any nonzero element in the range of σ_c . Since $f(c) \neq 0$, $f(c)$ is not in the kernel of σ_c . Thus, $f(c)$ is not a multiple of $p(x)$, and since $p(x)$ is irreducible, it follows that $f(c)$ and $p(x)$ are *relatively prime*. Therefore there are polynomials $s(x)$ and $t(x)$ such that $s(x)f(c) + t(x)p(x) = 1$. But then

$$s(c)f(c) + \underbrace{t(c)p(c)}_{=0} = 1$$

and therefore $s(c)$ is the multiplicative inverse of $f(c)$.

We have just shown that *the range of σ_c is a subfield of E* . Now, the range of σ_c is the set of all the elements $a(c)$, for all $a(x)$ in $F[x]$:

$$\text{Range } \sigma_c = \{a(c) : a(x) \in F[x]\}$$

We have just seen that $\text{range } \sigma_c$ is a field. In fact, it is the *smallest field containing F and c* : indeed, any other field containing F and c would inevitably contain every element of the form

$$a_0 + a_1c + \dots + a_nc^n \quad (a_0, \dots, a_n \in F)$$

in other words, would contain every element in the range of σ_c .

By the *smallest field containing F and c* we mean the field which contains F and c and is contained in any other field containing F and c . It is called the field *generated by F and c* , and is denoted by the important symbol

$$F(c)$$

Now, here is what we have, in a nutshell: σ_c is a homomorphism with domain $F[x]$, range $F(c)$, and kernel $J_c = \langle p(x) \rangle$. Thus, by the fundamental homomorphism theorem,

$$\boxed{F(c) \cong F[x] / \langle p(x) \rangle}$$

Finally, here is an interesting sidelight: if c and d are both roots of $p(x)$, where c and d are in E , then, by what we have just proved, $F(c)$ and $F(d)$ are both isomorphic to $F[x]/\langle p(x) \rangle$, and therefore isomorphic to each other:

If c and d are roots of the same irreducible polynomial $p(x)$ in $F[x]$, then $F(c) \cong F(d)$

In particular, this shows that, given F and c , $F(c)$ is unique up to isomorphism.

It is time now to recall our main objective: if $a(x)$ is a polynomial in $F[x]$ which has no roots in F , we wish to enlarge F to a field E which contains a root of $a(x)$. How can we manage this?

An observation is in order: finding extensions of F is not as easy as finding subfields of F . A subfield of F is a subset of an *existing* set: *it is there!* But an extension of F is *not yet there*. We must somehow build it around F .

Let $p(x)$ be an irreducible polynomial in $F[x]$. We have just seen that if F can be enlarged to a field E containing a root c of $p(x)$, then $F(c)$ is already what we are looking for: it is an extension of F containing a root of $p(x)$. Furthermore, $F(c)$ is isomorphic to $F[x]/\langle p(x) \rangle$. Thus, *the field extension we are searching for is precisely $F[x]/\langle p(x) \rangle$* . Our result is summarized in the next theorem.

Basic theorem of field extensions *Let F be a field and $a(x)$ a nonconstant polynomial in $F[x]$. There exists an extension field E of F and an element c in E such that c is a root of $a(x)$.*

PROOF: To begin with, $a(x)$ can be factored into irreducible polynomials in $F[x]$. If $p(x)$ is any nonconstant irreducible factor of $a(x)$, it is clearly sufficient to find an extension of F containing a root of $p(x)$, since such a root will also be a root of $a(x)$.

In [Exercise D4](#) of [Chapter 25](#), the reader was asked to supply the simple proof that, if $p(x)$ is irreducible in $F[x]$, then $\langle p(x) \rangle$ is a maximal ideal of $F[x]$. Furthermore, by the argument at the end of [Chapter 19](#), if $\langle p(x) \rangle$ is a maximal ideal of $F[x]$, then the quotient ring $F[x]/\langle p(x) \rangle$ is a field.

It remains only to prove that $F[x]/\langle p(x) \rangle$ is the desired field extension of F . When we write $J = \langle p(x) \rangle$, let us remember that every element of $F[x]/J$ is a coset of J . We will prove that $F[x]/J$ is an extension of F by identifying each element a in F with its coset $J + a$.

To be precise, define $h: F \rightarrow F[x]/J$ by $h(a) = J + a$. Note that h is the function which matches every a in F with its coset $J + a$ in $F[x]/J$. We will now show that h is an isomorphism.

By the familiar rules of coset addition and multiplication, h is a homomorphism. Now, every homomorphism between fields is injective. (This is true because the kernel of a homomorphism is an ideal, and a field has no nontrivial ideals.) Thus, h is an isomorphism between its domain and its range.

What is the range of h ? It consists of all the cosets $J + a$ where $a \in F$, that is, all the cosets of constant polynomials. (If a is in F , then a is a constant polynomial.) Thus, *F is isomorphic to the subfield of $F[x]/J$ containing all the cosets of constant polynomials*. This subfield is therefore an isomorphic copy of F , which may be identified with F , so $F[x]/J$ is an extension of F .

Finally, if $p(x) = a_0 + a_1x + \cdots + a_nx^n$, let us show that the coset $J + x$ is a root of $p(x)$ in $F[x]/J$. Of course, in $F[x]/J$, the coefficients are not actually $a_0, a_1, \dots, a_n$, but their cosets $J + a_0, J + a_1, \dots, J + a_n$. Writing

$$J + a_0 = \bar{a}_0, \dots, J + a_n = \bar{a}_n \quad \text{and} \quad J + x = \bar{x}$$

we must prove that

$$\bar{a}_0 + \bar{a}_1\bar{x} + \cdots + \bar{a}_n\bar{x}^n = J \quad (J \text{ is the zero coset})$$

Well,

$$\begin{aligned} \bar{a}_0 + \bar{a}_1\bar{x} + \cdots + \bar{a}_n\bar{x}^n &= (J + a_0) + (J + a_1)(J + x) + \cdots + (J + a_n)(J + x)^n \\ &= (J + a_0) + (J + a_1x) + \cdots + (J + a_nx^n) \\ &= J + p(x) \\ &= J \quad [\text{because } p(x) \in J] \end{aligned}$$

This completes the proof of the basic theorem of field extensions. Observe that we may use this theorem several times in succession to get the following:

Let $a(x)$ be a polynomial of degree n in $F[x]$, There is an extension field E of F which contains all n roots of $a(x)$.

EXERCISES

A. Recognizing Algebraic Elements

Example To show that $\sqrt{1+\sqrt{2}}$ is algebraic over $\mathbf{Q}$, one must find a polynomial $p(x) \in \mathbf{Q}[x]$ such that $\sqrt{1+\sqrt{2}}$ is a root of $p(x)$.

Let $a = \sqrt{1+\sqrt{2}}$; then $a^2 = 1 + \sqrt{2}$, $a^2 - 1 = \sqrt{2}$, and finally, $(a^2 - 1)^2 = 2$. Thus, a satisfies $p(x) = x^4 - 2x^2 - 1 = 0$.

1 Prove that each of the following numbers is algebraic over $\mathbf{Q}$:

- (a) i
- (b) $\sqrt{2}$
- (c) $2 + 3i$
- (d) $\sqrt{1+\sqrt[3]{2}}$
- #(e) $\sqrt{i-\sqrt{2}}$
- (f) $\sqrt{2} + \sqrt{3}$
- (g) $\sqrt[3]{2} + \sqrt[3]{4}$

2 Prove that each of the following numbers is algebraic over the given field:

- (a) $\sqrt{\pi}$ over $\mathbf{Q}(\pi)$
- (b) $\sqrt{\pi}$ over $\mathbf{Q}(\pi^2)$
- (c) $\pi^2 - 1$ over $\mathbf{Q}(\pi^3)$

NOTE: Recognizing a transcendental element is much more difficult, since it requires proving that the element cannot be a root of *any* polynomial over the given field. In recent times it has been proved, using sophisticated mathematical machinery, that π and e are transcendental over $\mathbf{Q}$.

B. Finding the Minimum Polynomial

1 Find the minimum polynomial of each of the following numbers over $\mathbf{Q}$. (Where appropriate, use the methods of [Chapter 26](#), [Exercises D](#), [E](#), and [F](#) to ensure that your polynomial is irreducible.)

- (a) $1 + 2i$
- (b) $1 + \sqrt{2}$
- (c) $1 + \sqrt{2}i$
- #(d) $\sqrt{2+\sqrt[3]{3}}$
- (e) $\sqrt{3} + \sqrt{5}$
- (f) $\sqrt{1+\sqrt{2}}$

2 Show that the minimum polynomial of $\sqrt{2} + i$ is

- (a) $x^2 - 2\sqrt{2}x + 3$ over $\mathbf{R}$
- (b) $x^4 - 2x^2 + 9$ over $\mathbf{Q}$
- (c) $x^2 - 2ix - 3$ over $\mathbf{Q}(i)$

3 Find the minimum polynomial of the following numbers over the indicated fields:

$$\sqrt{3} + i \quad \text{over } \mathbf{R}; \text{ over } \mathbf{Q}; \text{ over } \mathbf{Q}(i); \text{ over } \mathbf{Q}(\sqrt{3})$$

$\sqrt{i+\sqrt{2}}$ over $\mathbb{R}$; over $\mathbb{Q}(i)$; over $\mathbb{Q}(\sqrt{2})$; over $\mathbb{Q}$

4 For each of the following polynomials $p(x)$, find a number a such that $p(x)$ is the minimum polynomial of a over $\mathbb{Q}$:

- (a) $x^2 + 2x - 1$
- (b) $x^4 + 2x^2 - 1$
- (c) $x^4 - 10x^2 + 1$

5 Find a monic irreducible polynomial $p(x)$ such that $\mathbb{Q}[x]/\langle p(x) \rangle$ is isomorphic to

- (a) $\mathbb{Q}(\sqrt{2})$
- (b) $\mathbb{Q}(1 + \sqrt{2})$
- (c) $\mathbb{Q}(\sqrt{1 + \sqrt{2}})$

C. The Structure of Fields $F[x]/\langle p(x) \rangle$

Let $p(x)$ be an irreducible polynomial of degree n over F . Let c denote a root of $p(x)$ in some extension of F (as in the basic theorem on field extensions).

1 Prove: Every element in $F(c)$ can be written as $r(c)$, for some $r(x)$ of degree $< n$ in $F[x]$. [HINT: Given any element $t(c) \in F(c)$, use the division algorithm to divide $t(x)$ by $p(x)$.]

2 If $s(c) = t(c)$ in $F(c)$, where $s(x)$ and $t(x)$ have degree $< n$, prove that $s(x) = t(x)$.

3 Conclude from parts 1 and 2 that every element in $F(c)$ can be written *uniquely* as $r(c)$, with $\deg r(x) < n$.

4 Using part 3, explain why there are exactly four elements in $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$. List these four elements, and give their addition and multiplication tables. {HINT: Identify $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$ with $\mathbb{Z}_2(c)$, where c is a root of $x^2 + x + 1$. Write the elements of $\mathbb{Z}_2(c)$ as in part 3. When computing the multiplication table, use the fact that $c^2 + c + 1 = 0$.}

5 Describe $\mathbb{Z}_2[x]/\langle x^3 + x + 1 \rangle$, as in part 4.

6 Describe $\mathbb{Z}_3[x]/\langle x^3 + x^2 + 2 \rangle$, as in part 4.

D. Short Questions Relating of Field Extensions

Let F be any field.

Prove parts 1–5:

1 If c is algebraic over F , so are $c + 1$ and kc (where $k \in F$).

2 If $c \neq 0$ and c is algebraic over F , so is $1/c$.

3 If cd is algebraic over F , then c is algebraic over $F(d)$. If $c + d$ is algebraic over F , then c is algebraic over $F(d)$ (Assume $c \neq 0$ and $d \neq 0$.)

4 If the minimum polynomial of a over F is of degree 1, then $a \in F$, and conversely.

5 Suppose $F \subseteq K$ and $a \in K$. If $p(x)$ is a monic irreducible polynomial in $F[x]$, and $p(a) = 0$, then $p(x)$ is the minimum polynomial of a over F .

6 Name a field ($\neq \mathbb{R}$ or $\mathbb{C}$) which contains a root of $x^5 + 2x^3 + 4x^2 + 6$.

7 Prove: $\mathbb{Q}(1 + i) \cong \mathbb{Q}(1 - i)$. However, $\mathbb{Q}(\sqrt{2}) \not\cong \mathbb{Q}(\sqrt{3})$.

8 If $p(x)$ is irreducible and has degree 2, prove that $F[x]/\langle p(x) \rangle$ contains *both* roots of $p(x)$.

E. Simple Extensions

Recall the definition of $F(a)$. It is a field such that (i) $F \subseteq F(a)$; (ii) $a \in F(a)$; (iii) any field containing F and a contains $F(a)$.

Use this definition to prove parts 1–5, where $F \subseteq K$, $c \in F$, and $a \in K$:

- 1 $F(a) = F(a + c)$ and $F(a) = F(ca)$. (Assume $c \neq 0$.)
- 2 $F(a^2) \subseteq F(a)$ and $F(a + b) \subseteq F(a, b)$. [$F(a, b)$ is the field containing F , a , and b , and contained in any other field containing F , a and b .] Why are the reverse inclusions not necessarily true?
- 3 $a + c$ is a root of $p(x)$ iff a is a root of $p(x + c)$; ca is a root of $p(x)$ iff a is a root of $p(cx)$.
- 4 Let $p(x)$ be irreducible, and let a be a root of $p\{x + c\}$. Then

$$F[x]/\langle p(x + c) \rangle \cong F(a) \quad \text{and} \quad F[x]/\langle p(x) \rangle \cong F(a + c)$$

Conclude that $F[x]/\langle p(x + c) \rangle \cong F[x]/\langle p(x) \rangle$.

- 5 Let $p(x)$ be irreducible, and let a be a root of $p(cx)$. Then $F[x]/\langle p(cx) \rangle \cong F(a)$ and $F[x]/\langle p(x) \rangle \cong F(ca)$. Conclude that $F[x]/\langle p(cx) \rangle \cong F[x]/\langle p(x) \rangle$.

- 6 Use parts 4 and 5 to prove the following:

- (a) $\mathbb{Z}_{11}[x]/\langle x^2 + 1 \rangle \cong \mathbb{Z}_{11}[x]/\langle x^2 + x + 4 \rangle$.
- (b) If a is a root of $x^2 - 2$ and b is a root of $x^2 - 4x + 2$, then $\mathbb{Q}(a) \cong \mathbb{Q}(b)$.
- (c) If a is a root of $x^2 - 2$ and b is a root of $x^2 - \frac{1}{2}$, then $\mathbb{Q}(a) \cong \mathbb{Q}(b)$.

† F. Quadratic Extensions

If the minimum polynomial of a over F has degree 2, we call $F(a)$ a quadratic extension of F .

- 1 Prove that, if F is a field whose characteristic is $\neq 2$, any quadratic extension of F is of the form $F(\sqrt{a})$, for some $a \in F$ (HINT: Complete the square, and use [Exercise E4](#).)

Let F be a finite field, and F^* the multiplicative group of nonzero elements of F . Obviously $H = \{x^2: x \in F^*\}$ is a subgroup of F^* ; since every square x^2 in F^* is the square of only two different elements, namely $\pm x$, exactly half the elements of F^* are in H . Thus, H has exactly two cosets: H itself, containing all the squares, and aH (where $a \notin H$), containing all the nonsquares. If a and b are nonsquares, then by [Chapter 15, Theorem 5\(i\)](#),

$$ab^{-1} = \frac{a}{b} \in H$$

Thus: if a and b are nonsquares, a/b is a square. Use these remarks in the following:

- 2 Let F be a finite field. If $a, b \in F$, let $p(x) = x^2 - a$ and $q(x) = x^2 - b$ be irreducible in $F[x]$, and let $\sqrt{a}$ and $\sqrt{b}$ denote roots of $p(x)$ and $q(x)$ in an extension of F . Explain why a/b is a square, say $a/b = c^2$ for some $c \in F$. Prove that $\sqrt{b}$ is a root of $p(cx)$.
- 3 Use part 2 to prove that $F[x]/\langle p(cx) \rangle \cong F(\sqrt{b})$; then use [Exercise E5](#) to conclude that $F(\sqrt{a}) \cong F(\sqrt{b})$.
- 4 Use part 3 to prove: Any two quadratic extensions of a finite field are isomorphic.
- 5 If a and b are nonsquares in $\mathbb{R}$, a/b is a square (why?). Use the same argument as in part 4 to prove that any two simple extensions of $\mathbb{R}$ are isomorphic (hence isomorphic to $\mathbb{C}$).

G. Questions Relating to Transcendental Elements

Let F be a field, and let c be transcendental over F . Prove the following:

- 1 $\{a(c): a(x) \in F[x]\}$ is an integral domain isomorphic to $F[x]$.
- # 2 $F(c)$ is the field of quotients of $\{a(c): a(x) \in F[x]\}$, and is isomorphic to $F(x)$, the field of quotients of $F[x]$.
- 3 If c is transcendental over F , so are $c + 1$, kc (where $k \in F$ and $k \neq 0$), c^2 .
- 4 If c is transcendental over F , every element in $F(c)$ but not in F is transcendental over F .

† H. Common Factors of Two Polynomials: Over F and over Extensions of F

Let F be a field, and let $a(x), b(x) \in F[x]$. Prove the following:

- 1 If $a(x)$ and $b(x)$ have a common root c in some extension of F , they have a common factor of positive degree in $F[x]$. [Use the fact that $a(x), b(x) \in \ker \sigma_c$.]
- 2 If $a(x)$ and $b(x)$ are relatively prime in $F[x]$, they are relatively prime in $K[x]$, for any extension K of F . Conversely, if they are relatively prime in $K[x]$, then they are relatively prime in $F[x]$.

† I. Derivatives and Their Properties

Let $a(x) = a_0 + a_1x + \dots + a_nx^n \in F[x]$. The *derivative* of $a(x)$ is the following polynomial $a'(x) \in F[x]$:

$$a'(x) = a_1 + 2a_2x + \dots + na_nx^{n-1}$$

(This is the same as the derivative of a polynomial in calculus.) We now prove the analogs of the formal rules of differentiation, familiar from calculus.

Let $a(x), b(x) \in F[x]$, and let $k \in F$.

Prove parts 1–4:

- 1 $[a(x) + b(x)]' = a'(x) + b'(x)$
- 2 $[a(x)b(x)]' = a'(x)b(x) + a(x)b'(x)$
- 3 $[ka(x)]' = ka'(x)$
- 4 If F has characteristic 0 and $a'(x) = 0$, then $a(x)$ is a constant polynomial. Why is this conclusion not necessarily true if F has characteristic $p \neq 0$?
- 5 Find the derivative of the following polynomials in $\mathbb{Z}_5[x]$:

$$x^6 + 2x^3 + x + 1 \quad x^5 + 3x^2 + 1 \quad x^{15} + 3x^{10} + 4x^5 + 1$$

- 6 If F has characteristic $p \neq 0$, and $a'(x) = 0$, prove that the only nonzero terms of $a(x)$ are of the form $a_{mp}x^{mp}$ for some m . [That is, $a(x)$ is a polynomial in powers of x^p .]

† J. Multiple Roots

Suppose $a(x) \in F[x]$, and K is an extension of F . An element $c \in K$ is called a multiple root of $a(x)$ if $(x - c)^m | a(x)$ for some $m > 1$. It is often important to know if all the roots of a polynomial are different, or

not. We now consider a method for determining whether an arbitrary polynomial $a(x) \in F[x]$ has multiple roots in any extension of F .

Let K be any field containing all the roots of $a(x)$. Suppose $a(x)$ has a multiple root c .

1 Prove that $a(x) = (x - c)^2 q(x) \in K[x]$.

2 Compute $a'(x)$, using part 1.

3 Show that $x - c$ is a common factor of $a(x)$ and $a'(x)$. Use Exercise 1 to conclude that $a(x)$ and $a'(x)$ have a common factor of degree > 1 in $F[x]$.

Thus, if $a(x)$ has a multiple root, then $a(x)$ and $a'(x)$ have a common factor in $F[x]$. To prove the converse, suppose $a(x)$ has *no* multiple roots. Then $a(x)$ can be factored as $a(x) = (x - c_1) \cdots (x - c_n)$ where $c_1, \dots, c_n$ are all different.

4 Explain why $a'(x)$ is a sum of terms of the form

$$(x - c_1) \cdots (x - c_{i-1})(x - c_{i+1}) \cdots (x - c_n)$$

5 Using part 4, explain why none of the roots $c_1, \dots, c_n$ of $a(x)$ are roots of $a'(x)$.

6 Conclude that $a(x)$ and $a'(x)$ have no common factor of degree > 1 in $F[x]$.

This important result is stated as follows: *A polynomial $a(x)$ in $F[x]$ has a multiple root iff $a(x)$ and $a'(x)$ have a common factor of degree > 1 in $F[x]$.*

7 Show that each of the following polynomials has *no* multiple roots in any extension of its field of coefficients:

$$x^3 - 7x^2 + 8 \in \mathbf{Q}[x] \quad x^2 + x + 1 \in \mathbf{Z}_5[x] \quad x^{100} - 1 \in \mathbf{Z}_7[x]$$

The preceding example is most interesting: it shows that there are 100 *different* hundredth roots of 1 over $\mathbf{Z}_7$. (The roots ± 1 are in $\mathbf{Z}_7$, while the remaining 98 roots are in extensions of $\mathbf{Z}_7$.) Corresponding results hold for most other fields.

CHAPTER TWENTY-EIGHT

VECTOR SPACES

Many physical quantities, such as length, area, weight, and temperature, are completely described by a single real number. On the other hand, many other quantities arising in scientific measurement and everyday reckoning are best described by a combination of *several* numbers. For example, a point in space is specified by giving its three coordinates with respect to an *xyz* coordinate system.

Here is an example of a different kind: A store handles 100 items; its monthly inventory is a sequence of 100 numbers $(a_1, a_2, \dots, a_{100})$ specifying the quantities of each of the 100 items currently in stock. Such a sequence of numbers is usually called a *vector*. When the store is restocked, a vector is *added* to the current inventory vector. At the end of a good month of sales, a vector is *subtracted*.

As this example shows, it is natural to add vectors by adding corresponding components, and subtract vectors by subtracting corresponding components. If the store manager in the preceding example decided to double inventory, each component of the inventory vector would be multiplied by 2. This shows that a natural way of multiplying a vector by a real number k is to multiply each component by k . This kind of multiplication is commonly called *scalar multiplication*.

Historically, as the use of vectors became widespread and they came to be an indispensable tool of science, vector algebra grew to be one of the major branches of mathematics. Today it forms the basis for much of advanced calculus, the theory and practice of differential equations, statistics, and vast areas of applied mathematics. Scientific computation is enormously simplified by vector methods; for example, 3, or 300, or 3000 individual readings of scientific instruments can be expressed as a single vector.

In any branch of mathematics it is elegant and desirable (but not always possible) to find a *simple* list of axioms from which all the required theorems may be proved. In the specific case of vector algebra, we wish to select as axioms only those particular properties of vectors which are absolutely necessary for proving further properties of vectors. And we must select a sufficiently complete list of axioms so that, by using them and them alone, we can prove all the properties of vectors needed in mathematics.

A delightfully simple list of axioms is available for vector algebra. The remarkable fact about this axiom system is that, although we conceive of vectors as finite sequences $(a_1, a_2, \dots, a_n)$ of numbers,

nothing in the axioms actually requires them to be such sequences! Instead, vectors are treated simply as elements in a set, satisfying certain equations. Here is our basic definition:

A *vector space* over a field F is a set V , with two operations $+$ and $\cdot$ called *vector addition* and *scalar multiplication*, such that

1. V with vector addition is an abelian group.
2. For any $k \in F$ and $\mathbf{a} \in V$, the scalar product $k\mathbf{a}$ is an element of V , subject to the following conditions:
for all $k, l \in F$ and $\mathbf{a}, \mathbf{b} \in V$,
 - (a) $k(\mathbf{a} + \mathbf{b}) = k\mathbf{a} + k\mathbf{b}$,
 - (b) $(k + l)\mathbf{a} = k\mathbf{a} + l\mathbf{a}$,
 - (c) $k(l\mathbf{a}) = (kl)\mathbf{a}$,
 - (d) $1\mathbf{a} = \mathbf{a}$.

The elements of V are called *vectors* and the elements of the field F are called *scalars*,

In the following exposition the field F will not be specifically referred to unless the context requires it. For notational clarity, vectors will be written in **bold type** and scalars in *italics*.

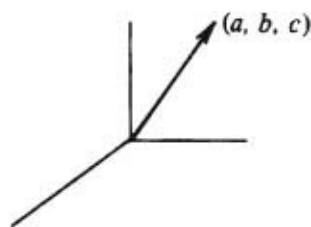
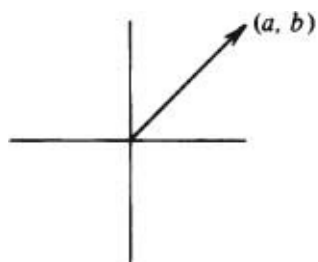
The traditional example of a vector space is the set $\mathbb{R}^n$ of all n -tuples of real numbers, $(a_1, a_2, \dots, a_n)$, with the operations

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

$$\text{and} \quad k(a_1, a_2, \dots, a_n) = (ka_1, ka_2, \dots, ka_n)$$

For example, $\mathbb{R}^2$ is the set of all two-dimensional vectors (a, b) , while $\mathbb{R}^3$ is the set of all vectors (a, b, c) in euclidean space. (See the figure on the next page.)

However, these are not the only vector spaces! Our definition of vector space is so very simple that many other things, quite different in appearance from the traditional vector spaces, satisfy the conditions of our definition and are therefore, legitimately, vector spaces.



For example, $\mathcal{FR}$, you may recall, is the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. We define the sum $f + g$ of two functions by the rule

$$[f + g](x) = f(x) + g(x)$$

and we define the product af , of a real number a and a function f , by

$$[af](x) = af(x)$$

It is very easy to verify that $\mathcal{FR}$, with these operations, satisfies all the conditions needed in order to be a vector space over the field $\mathbb{R}$.

As another example, let $\mathcal{PL}$ denote the set of all polynomials with real coefficients. Polynomials are

added as usual, and scalar multiplication is defined by

$$k(a_0 + a_1x + \cdots + a_nx^n) = (ka_0) + (ka_1)x + \cdots + (ka_n)x^n$$

Again, it is not hard to see that $\mathcal{P}\ell$ is a vector space over $\mathbb{R}$.

Let V be a vector space. Since V with addition alone is an abelian group, there is a zero element in V called the *zero vector*, written as $\mathbf{0}$. Every vector $\mathbf{a}$ in V has a negative, written as $-\mathbf{a}$. Finally, since V with vector addition is an abelian group, it satisfies the following conditions which are true in all abelian groups:

$$\mathbf{a} + \mathbf{b} = \mathbf{a} + \mathbf{c} \quad \text{implies} \quad \mathbf{b} = \mathbf{c} \quad (1)$$

$$\mathbf{a} + \mathbf{b} = \mathbf{0} \quad \text{implies} \quad \mathbf{a} = -\mathbf{b} \quad \text{and} \quad \mathbf{b} = -\mathbf{a} \quad (2)$$

$$-(\mathbf{a} + \mathbf{b}) = (-\mathbf{a}) + (-\mathbf{b}) \quad \text{and} \quad -(-\mathbf{a}) = \mathbf{a} \quad (3)$$

There are simple, obvious rules for multiplication by zero and by negative scalars. They are contained in the next theorem.

Theorem 1 *If V is a vector space, then:*

- (i) $0\mathbf{a} = \mathbf{0}$, for every $\mathbf{a} \in V$
- (ii) $k\mathbf{0} = \mathbf{0}$, for every scalar k .
- (iii) If $k\mathbf{a} = \mathbf{0}$, then $k = 0$ or $\mathbf{a} = \mathbf{0}$.
- (iv) $(-1)\mathbf{a} = -\mathbf{a}$ for every $\mathbf{a} \in V$.

To prove Rule (i), we observe that

$$0\mathbf{a} = (0 + 0)\mathbf{a} = 0\mathbf{a} + 0\mathbf{a}$$

hence $\mathbf{0} + 0\mathbf{a} = 0\mathbf{a} + 0\mathbf{a}$. It follows by Condition (1) that $\mathbf{0} = 0\mathbf{a}$.

Rule (ii) is proved similarly. As for Rule (iii), if $k = 0$, we are done. If $k \neq 0$, we may multiply $k\mathbf{a} = \mathbf{0}$ by $1/k$ to get $\mathbf{a} = \mathbf{0}$. Finally, for Rule (iv), we have

$$\mathbf{a} + (-1)\mathbf{a} = 1\mathbf{a} + (-1)\mathbf{a} = (1 + (-1))\mathbf{a} = 0\mathbf{a} = \mathbf{0}$$

so by Condition (2), $(-1)\mathbf{a} = -\mathbf{a}$.

Let V be a vector space, and $U \subseteq V$. We say that U is *closed with respect to scalar multiplication* if $k\mathbf{a} \in U$ for every scalar k and every $\mathbf{a} \in U$. We call U a *subspace* of V if U is *closed with respect to addition and scalar multiplication*. It is easy to see that if V is a vector space over the field F , and U is a subspace of V , then U is a vector space over the same field F .

If $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are in V and $k_1, k_2, \dots, k_n$ are scalars, then the vector

$$k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + \cdots + k_n\mathbf{a}_n$$

is called a *linear combination* of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. The set of all the linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ is a subspace of V . (This fact is exceedingly easy to verify.)

If U is the subspace consisting of all the linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, we call U the

subspace *spanned* by $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. An equivalent way of saying the same thing is as follows: a space (or subspace) U is spanned by $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ iff every vector in U is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$.

If U is spanned by $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, we also say that $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ *span* U .

Let $S = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ be a set of distinct vectors in a vector space V . Then S is said to be *linearly dependent* if there are scalars $k_1, \dots, k_n$, *not all zero*, such that

$$k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2 + \dots + k_n \mathbf{a}_n = \mathbf{0} \quad (4)$$

Obviously this is the same as saying that at least one of the vectors in S is a linear combination of the remaining ones. [Solve for any vector $\mathbf{a}_i$ in Equation (4) having a nonzero coefficient.]

If $S = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ is not linearly dependent, then it is *linearly independent*. That is, S is linearly independent iff

$$k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2 + \dots + k_n \mathbf{a}_n = \mathbf{0} \quad \text{implies} \quad k_1 = k_2 = \dots = k_n = 0$$

This is the same as saying that *no* vector in S is equal to a linear combination of the other vectors in S .

It is obvious from these definitions that any set of vectors containing the zero vector is linearly dependent. Furthermore, the set $\{\mathbf{a}\}$, containing a single nonzero vector $\mathbf{a}$, is linearly independent.

The next two lemmas, although very easy and at first glance rather trite, are used to prove the most fundamental theorems of this subject.

Lemma 1 *If $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ is linearly dependent, then some $\mathbf{a}_i$ is a linear combination of the preceding ones, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}$.*

PROOF: Indeed, if $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ is linearly dependent, then $k_1 \mathbf{a}_1 + \dots + k_n \mathbf{a}_n = \mathbf{0}$ for coefficients $k_1, k_2, \dots, k_n$ which are not all zero. If k_i is the last nonzero coefficient among them, then $k_1 \mathbf{a}_1 + \dots + k_i \mathbf{a}_i = \mathbf{0}$, and this equation can be used to solve for $\mathbf{a}_i$ in terms of $\mathbf{a}_1, \dots, \mathbf{a}_{i-1}$. ■

Let $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n\}$ denote the set $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ after removal of $\mathbf{a}_i$.

Lemma 2 *If $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ spans V , and $\mathbf{a}_i$ is a linear combination of preceding vectors, then $\{\mathbf{a}_1, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n\}$ still spans V .*

PROOF: Our assumption is that $\mathbf{a}_i = k_1 \mathbf{a}_1 + \dots + k_{i-1} \mathbf{a}_{i-1}$ for some scalars $k_1, \dots, k_{i-1}$. Since every vector $\mathbf{b} \in V$ is a linear combination

$$\mathbf{b} = l_1 \mathbf{a}_1 + \dots + l_i \mathbf{a}_i + \dots + l_n \mathbf{a}_n$$

it can also be written as a linear combination

$$\mathbf{b} = l_1 \mathbf{a}_1 + \dots + l_i (k_1 \mathbf{a}_1 + \dots + k_{i-1} \mathbf{a}_{i-1}) + \dots + l_n \mathbf{a}_n$$

in which $\mathbf{a}_i$ does not figure. ■

A set of vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ in V is called a *basis* of V if it is *linearly independent* and *spans* V .

For example, the vectors $\varepsilon_1 = (1, 0, 0)$, $\varepsilon_2 = (0, 1, 0)$, and $\varepsilon_3 = (0, 0, 1)$ form a basis of $\mathbb{R}^3$. They are

linearly independent because, obviously, no vector in $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ is equal to a linear combination of preceding ones. [Any linear combination of ε_1 and ε_2 is of the form $a\varepsilon_1 + b\varepsilon_2 = (a, b, 0)$, whereas ε_3 is not of this form; similarly, any linear combination of ε_1 alone is of the form $a\varepsilon_1 = (a, 0, 0)$, and ε_2 is not of that form.] The vectors $\varepsilon_1, \varepsilon_2, \varepsilon_3$ span $\mathbb{R}^3$ because any vector (a, b, c) in $\mathbb{R}^3$ can be written as $(a, b, c) = a\varepsilon_1 + b\varepsilon_2 + c\varepsilon_3$.

Actually, $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ is not the only basis of $\mathbb{R}^3$. Another basis of $\mathbb{R}^3$ consists of the vectors $(1, 2, 3)$, $(1, 0, 2)$, and $(3, 2, 1)$; in fact, there are infinitely many different bases of $\mathbb{R}^3$. Nevertheless, all bases of $\mathbb{R}^3$ have one thing in common: they contain exactly three vectors! This is a consequence of our next theorem:

Theorem 2 *Any two bases of a vector space V have the same number of elements.*

PROOF: Suppose, on the contrary, that V has a basis $A = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ where $m \neq n$. To be specific, suppose $n < m$. From this assumption we will derive a contradiction.

Put the vector $\mathbf{b}_1$ in the set A , so A now contains $\{\mathbf{b}_1, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$. This set is linearly dependent because $\mathbf{b}_1$ is a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$. But then, by Lemma 1, some $\mathbf{a}_i$ is a linear combination of preceding vectors. By Lemma 2 we may expel this $\mathbf{a}_i$, and the remaining set $\{\mathbf{b}_1, \mathbf{a}_1, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n\}$ still spans V .

Repeat this argument a second time by putting $\mathbf{b}_2$ in A , so A now contains $\{\mathbf{b}_2, \mathbf{b}_1, \mathbf{a}_1, \mathbf{a}_2, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n\}$. This set is linearly dependent because $\{\mathbf{b}_1, \mathbf{a}_1, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n\}$ spans V and therefore $\mathbf{b}_2$ is a linear combination of $\mathbf{b}_1, \mathbf{a}_1, \dots, \cancel{\mathbf{a}_i}, \dots, \mathbf{a}_n$. By Lemma 1, some $\mathbf{a}_j$ is a linear combination of preceding vectors in A , so by Lemma 2 we may remove $\mathbf{a}_j$, and $\{\mathbf{b}_2, \mathbf{b}_1, \mathbf{a}_1, \mathbf{a}_2, \dots, \cancel{\mathbf{a}_i}, \dots, \cancel{\mathbf{a}_j}, \mathbf{a}_n\}$ still spans V .

This argument is repeated n times. Each time, a vector from B is put into A and a vector $\mathbf{a}_k$ is removed. At the end of the n th repetition, A contains only $\mathbf{b}_1, \dots, \mathbf{b}_n$, and $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ still spans V . But this is impossible because it implies that $\mathbf{b}_{n+1}$ is a linear combination of $\mathbf{b}_1, \dots, \mathbf{b}_n$, whereas in fact, $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n, \dots, \mathbf{b}_m\}$ is linearly independent!

This contradiction proves that any two bases of V must contain *the same number of elements* ! ■

If V has a basis $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$, we call V a *finite-dimensional vector space* and say that V is of *dimension n* . In that case, by Theorem 2 every basis of V has exactly n elements.

In the sequel we consider only finite-dimensional vector spaces. The next two lemmas are quite interesting. The first one states that if $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ spans V , there is a way of removing vectors from this set, one by one, until we are left with an *independent* set which still spans V .

Lemma 3 *If the set $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ spans V , it contains a basis of V .*

PROOF: If $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ is an independent set, it is a basis, and we are done. If not, some $\mathbf{a}_1$ is a linear combination of preceding ones, so $\{\mathbf{a}_1, \dots, \cancel{\mathbf{a}_1}, \dots, \mathbf{a}_m\}$ still spans V . Repeating this process, we discard vectors one by one from $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ and, each time, the remaining vectors still span V . We keep doing this until the remaining set is independent. (In the worst case, this will happen when only one vector is left.) ■

The next lemma asserts that if $\{\mathbf{a}_1, \dots, \mathbf{a}_s\}$ is an independent set of vectors in V , there is a way of

adding vectors to this set so as to get a basis of V .

Lemma 4 *If the set $\{\mathbf{a}_1, \dots, \mathbf{a}_s\}$ is linearly independent, it can be extended to a basis of V .*

PROOF: If $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is any basis of V , then $\{\mathbf{a}_1, \dots, \mathbf{a}_s, \mathbf{b}_1, \dots, \mathbf{b}_n\}$ spans V . By the proof of [Lemma 3](#), we may discard vectors from this set until we get a basis of V . Note that we never discard any $\mathbf{a}_i$, because, by hypothesis, $\mathbf{a}_i$ is not a linear combination of preceding vectors. ■

The next theorem is an immediate consequence of [Lemmas 3](#) and [4](#).

Theorem 3 *Let V have dimension n . If $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is an independent set, it is already a basis of V . If $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ spans V , it is already a basis of V .*

If $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is a basis of V , then every vector $\mathbf{c}$ in V has a *unique* expression $\mathbf{c} = k_1\mathbf{a}_1 + \dots + k_n\mathbf{a}_n$ as a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$. Indeed, if

$$\mathbf{c} = k_1 \mathbf{a}_1 + \dots + k_n \mathbf{a}_n = l_1 \mathbf{a}_1 + \dots + l_n \mathbf{a}_n$$

then

$$(k_1 - l_1)\mathbf{a}_1 + \dots + (k_n - l_n)\mathbf{a}_n = \mathbf{0}$$

hence

$$k_1 - l_1 = \dots = k_n - l_n = 0$$

so $k_1 = l_1, \dots, k_n = l_n$. If $\mathbf{c} = k_1\mathbf{a}_1 + \dots + k_n\mathbf{a}_n$, the coefficients $k_1, \dots, k_n$ are called the *coordinates* of $\mathbf{c}$ with respect to the basis $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. It is then convenient to represent $\mathbf{c}$ as the n -tuple

$$\mathbf{c} = (k_1, \dots, k_n)$$

If U and V are vector spaces over a field F , a function $h : U \rightarrow V$ is a *homomorphism* if it satisfies the following two conditions:

$$h(a + b) = h(a) + h(b)$$

and

$$h(ka) = kh(a)$$

A homomorphism of vector spaces is also called a *linear transformation*.

If $h : U \rightarrow V$ is a linear transformation, its kernel [that is, the set of all $\mathbf{a} \in U$ such that $h(\mathbf{a}) = \mathbf{0}$] is a subspace of U , called the *null space* of h . Homomorphisms of vector spaces behave very much like homomorphisms of groups and rings. Their properties are presented in the exercises.

EXERCISES

A. Examples of Vector Spaces

- 1 Prove that $\mathbb{R}^n$, as defined on page 283, satisfies all the conditions for being a vector space over $\mathbb{R}$.
- 2 Prove that $\mathcal{F}(\mathbb{R})$, as defined on page 284, is a vector space over $\mathbb{R}$.
- 3 Prove that $\mathcal{P}\ell$, as defined on page 284, is a vector space over $\mathbb{R}$.
- 4 Prove that $\mathcal{M}_2(\mathbb{R})$, the set of all 2×2 matrices of real numbers, with matrix addition and the scalar multiplication

$$k \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix}$$

is a vector space over $\mathbb{R}$.

B. Examples of Subspaces

- # 1 Prove that $\{(a, b, c) : 2a - 3b + c = 0\}$ is a subspace of $\mathbb{R}^3$.
- 2 Prove that the set of all $(x, y, z) \in \mathbb{R}^3$ which satisfy the pair of equations $ax + by + c = 0$, $dx + ey + f = 0$ is a subspace of $\mathbb{R}^3$.
- 3 Prove that $\{f : f(1) = 0\}$ is a subspace of $\mathcal{F}(\mathbb{R})$.
- 4 Prove that $\{f : f \text{ is a constant on the interval } [0, 1]\}$ is a subspace of $\mathcal{F}(\mathbb{R})$.
- 5 Prove that the set of all even functions [that is, functions f such that $f(x) = f(-x)$] is a subspace of $\mathcal{F}(\mathbb{R})$. Is the same true for the set of all the odd functions [that is, functions f such that $f(-x) = -f(x)$]?
- 6 Prove that the set of all polynomials of degree $\leq n$ is a subspace of $\mathcal{P}\ell$

C. Examples of Linear Independence and Bases

- 1 Prove that $\{(0, 0, 0, 1), (0, 0, 1, 1), (0, 1, 1, 1), (1, 1, 1, 1)\}$ is a basis of $\mathbb{R}^4$.
- 2 If $\mathbf{a} = (1, 2, 3, 4)$ and $\mathbf{b} = (4, 3, 2, 1)$, explain why $\{\mathbf{a}, \mathbf{b}\}$ may be extended to a basis of $\mathbb{R}^4$. Then find a basis of $\mathbb{R}^4$ which includes $\mathbf{a}$ and $\mathbf{b}$.
- 3 Let A be the set of eight vectors (x, y, z) where $x, y, z = 1, 2$. Prove that A spans $\mathbb{R}^3$, and find a subset of A which is a basis of $\mathbb{R}^3$.
- 4 If $\mathcal{P}\ell_n$ is the subspace of $\mathcal{P}\ell$ consisting of all polynomials of degree $\leq n$, prove that $\{1, x, x^2, \dots, x^n\}$ is a basis of $\mathcal{P}\ell_n$. Then find another basis of $\mathcal{P}\ell_n$.
- 5 Find a basis for each of the following subspaces of $\mathbb{R}^3$:
 # (a) $S_1 = \{(x, y, z) : 3x - 2y + z = 0\}$ (b) $S_2 = \{(x, y, z) : x + y - z = 0 \text{ and } 2x - y + z = 0\}$
- 6 Find a basis for the subspace of $\mathbb{R}^3$ spanned by the set of vectors (x, y, z) such that $x^2 + y^2 + z^2 = 1$.
- 7 Let U be the subspace of $\mathcal{F}(\mathbb{R})$ spanned by $\{\cos^2 x, \sin^2 x, \cos 2x\}$. Find the dimension of U , and then find a basis of U .
- 8 Find a basis for the subspace of $\mathcal{P}\ell$ spanned by

$$\{x^3 + x^2 + x + 1, x^2 + 1, x^3 - x^2 + x - 1, x^2 - 1\}$$

D. Properties of Subspaces and Bases

Let V be a finite-dimensional vector space. Let $\dim V$ designate the dimension of V . Prove each of the following:

- 1 If U is a subspace of V , then $\dim U \leq \dim V$.
- 2 If U is a subspace of V , and $\dim U = \dim V$, then $U = V$.
- 3 Any set of vectors containing $\mathbf{0}$ is linearly dependent.
- 4 The set $\{\mathbf{a}\}$, containing only one nonzero vector $\mathbf{a}$, is linearly independent.
- 5 Any subset of an independent set is independent. Any set of vectors containing a dependent set is dependent.
- # 6 If $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is linearly independent, so is $\{\mathbf{a} + \mathbf{b}, \mathbf{b} + \mathbf{c}, \mathbf{a} + \mathbf{c}\}$.
- 7 If $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is a basis of V , so is $\{k_1\mathbf{a}_1, \dots, k_n\mathbf{a}_n\}$ for any nonzero scalars
- 8 The space spanned by $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is the same as the space spanned by $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ iff each $\mathbf{a}_i$ is a linear combination of $\mathbf{b}_1, \dots, \mathbf{b}_m$, and each $\mathbf{b}_j$ is a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$.

E. Properties of Linear Transformations

Let U and V be finite-dimensional vector spaces over a field F , and let $h : U \rightarrow V$ be a linear transformation. Prove parts 1–3:

- 1 The kernel of h is a subspace of U . (It is called the *null space* of h .)
- 2 The range of h is a subspace of V . (It is called the *range space* of h .)
- 3 h is injective iff the null space of h is equal to $\{\mathbf{0}\}$.

Let N be the null space of h , and $\mathcal{R}$ the range space of h . Let $\{\mathbf{a}_1, \dots, \mathbf{a}_r\}$ be a basis of N . Extend it to a basis $\{\mathbf{a}_1, \dots, \mathbf{a}_r, \dots, \mathbf{a}_n\}$ of U .

Prove parts 4–6:

- 4 Every vector $\mathbf{b} \in \mathcal{R}$ is a linear combination of $h(\mathbf{a}_{r+1}), \dots, h(\mathbf{a}_n)$.

5 $\{h(\mathbf{a}_{r+1}), \dots, h(\mathbf{a}_n)\}$ is linearly independent.

- 6 The dimension of $\mathcal{R}$ is $n - r$.

7 Conclude as follows: for any linear transformation h , $\dim(\text{domain } h) = \dim(\text{null space of } h) + \dim(\text{range space of } h)$.

- 8 Let U and V have the same dimension n . Use part 7 to prove that h is injective iff h is surjective.

F. Isomorphism of Vector Spaces

Let U and V be vector spaces over the field F , with $\dim U = n$ and $\dim V = m$. Let $h : U \rightarrow V$ be a homomorphism.

Prove the following:

- 1 Let h be injective. If $\{\mathbf{a}_1, \dots, \mathbf{a}_r\}$ is a linearly independent subset of U , then $\{h(\mathbf{a}_1), \dots, h(\mathbf{a}_r)\}$ is a linearly independent subset of V .

2 h is injective iff $\dim U = \dim h(U)$.

- 3 Suppose $\dim U = \dim V$; h is an isomorphism (that is, a bijective homomorphism) iff h is injective iff h is surjective.

- 4 Any n -dimensional vector space V over F is isomorphic to the space F^n of all n -tuples of elements of F .

†G. Sums of Vector Spaces

Let T and U be subspaces of V . The *sum* of T and U , denoted by $T + U$, is the set of all vectors $\mathbf{a} + \mathbf{b}$, where $\mathbf{a} \in T$ and $\mathbf{b} \in U$.

1 Prove that $T + U$ and $T \cap U$ are subspaces of V .

V is said to be the *direct sum* of T and U if $V = T + U$ and $T \cap U = \{\mathbf{0}\}$. In that case, we write $V = T \oplus U$.

2 Prove: $V = T \oplus U$ iff every vector $\mathbf{c} \in V$ can be written, in a unique manner, as a sum $\mathbf{c} = \mathbf{a} + \mathbf{b}$ where $\mathbf{a} \in T$ and $\mathbf{b} \in U$.

3 Let T be a k -dimensional subspace of an n -dimensional space V . Prove that an $(n - k)$ -dimensional subspace U exists such that $V = T \oplus U$.

4 If T and U are arbitrary subspaces of V , prove that

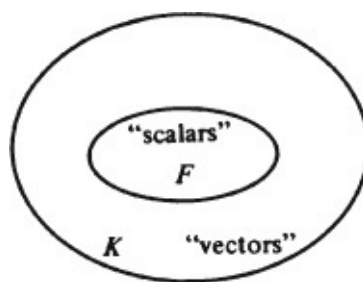
$$\dim(T + U) = \dim T + \dim U - \dim(T \cap U)$$

CHAPTER TWENTY-NINE

DEGREES OF FIELD EXTENSIONS

In this chapter we will see how the machinery of vector spaces can be applied to the study of field extensions.

Let F and K be fields. If K is an extension of F , we may regard K as being a vector space over F . We may treat the elements in K as “vectors”



and the elements in F as “scalars.” That is, when we add elements in K , we think of it as vector addition; when we add and multiply elements in F , we think of this as addition and multiplication of scalars; and finally, when we multiply an element of F by an element of K , we think of it as scalar multiplication.

We will be especially interested in the case where the resulting vector space is of finite dimension. If K , as a vector space over F , is of finite dimension, we call K a *finite extension* of F . If the dimension of the vector space K is n , we say that K is an *extension of degree n* over F . This is symbolized by writing

$$[K : F] = n$$

which should be read, “the degree of K over F is equal to n .”

Let us recall that $F(c)$ denotes the smallest field which contains F and c . This means that $F(c)$ contains F and c , and that any other field K containing F and c must contain $F(c)$. We saw in [Chapter 27](#) that if c is algebraic over F , then $F(c)$ consists of all the elements of the form $a(c)$, for all $a(x)$ in $F[x]$. Since $F(c)$ is an extension of F , we may regard it as a vector space over F . Is $F(c)$ a finite extension of

F ?

Well, let c be algebraic over F , and let $p(x)$ be the minimum polynomial of c over F . [That is, $p(x)$ is the monic polynomial of lowest degree having c as a root.] Let the degree of the polynomial $p(x)$ be equal to n . It turns out, then, that the n elements

$$1, c, c^2, \dots, c^{n-1}$$

are linearly independent and span $F(c)$. We will prove this fact in a moment, but meanwhile let us record what it means. It means that the set of n "vectors" $\{1, c, c^2, \dots, c^{n-1}\}$ is a basis of $F(c)$; hence $F(c)$ is a vector space of dimension n over the field F . This may be summed up concisely as follows:

Theorem 1 *The degree of $F(c)$ over F is equal to the degree of the minimum polynomial of c over F .*

PROOF: It remains only to show that the n elements $1, c, \dots, c^{n-1}$ span $F(c)$ and are linearly independent. Well, if $a(x)$ is any element of $F(c)$, use the division algorithm to divide $a(x)$ by $p(x)$:

$$a(x) = p(x)q(x) + r(x) \quad \text{where } \deg r(x) \leq n-1$$

Therefore,

$$a(c) = \underbrace{p(c)q(c)}_{=0} + r(c) = 0 + r(c) = r(c)$$

This shows that every element of $F(c)$ is of the form $r(c)$ where $r(x)$ has degree $n-1$ or less. Thus, every element of $F(c)$ can be written in the form

$$a_0 + a_1c + \dots + a_{n-1}c^{n-1}$$

which is a linear combination of $1, c, c^2, \dots, c^{n-1}$.

Finally, to prove that $1, c, c^2, \dots, c^{n-1}$ are linearly independent, suppose that $a_0 + a_1c + \dots + a_{n-1}c^{n-1} = 0$. If the coefficients $a_0, a_1, \dots, a_{n-1}$ were not all zero, c would be the root of a nonzero polynomial of degree $n-1$ or less, which is impossible because the minimum polynomial of c over F has degree n . Thus, $a_0 = a_1 = \dots = a_{n-1} = 0$. ■

For example, let us look at $\mathbb{Q}(\sqrt{2})$: the number $\sqrt{2}$ is not a root of any monic polynomial of degree 1 over $\mathbb{Q}$. For such a polynomial would

have to be $x - \sqrt{2}$, and the latter is not in $\mathbb{Q}[x]$ because $\sqrt{2}$ is irrational. However, $\sqrt{2}$ is a root of $x^2 - 2$, which is therefore the minimum polynomial of $\sqrt{2}$ over $\mathbb{Q}$, and which has degree 2. Thus,

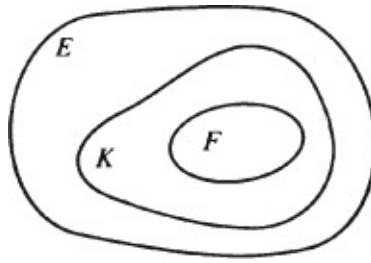
$$[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$$

In particular, every element in $\mathbb{Q}(\sqrt{2})$ is therefore a linear combination of 1 and $\sqrt{2}$, that is, a number of the form $a + b\sqrt{2}$ where $a, b \in \mathbb{Q}$.

As another example, i is a root of the irreducible polynomial $x^2 + 1$ over $\mathbb{R}$ in $\mathbb{R}[x]$. Therefore $x^2 + 1$ is the minimum polynomial of i over $\mathbb{R}$. $x^2 + 1$ has degree 2, so $[\mathbb{R}(i) : \mathbb{R}] = 2$. Thus, $\mathbb{R}(i)$ consists of all the

linear combinations of 1 and i with real coefficients, that is, all the $a + bi$ where $a, b \in \mathbb{R}$. Clearly then, $\mathbb{R}(i) = \mathbb{C}$, so the degree of $\mathbb{C}$ over $\mathbb{R}$ is equal to 2.

In the sequel we will often encounter the following situation: E is a finite extension of K , where K is a finite extension of F . If we know the



degree of E over K and the degree of K over F , can we determine the degree of E over F ? This is a question of major importance! Fortunately, it has an easy answer, based on the following lemma:

Lemma Let $a_1, a_2, \dots, a_m$ be a basis of the vector space K over F , and let $b_1, b_2, \dots, b_n$ be a basis of the vector space E over K . Then the set of mn products $\{a_i b_j\}$ is a basis of the vector space E over the field F .

PROOF: To prove that the set $\{a_i b_j\}$ spans E , note that each element c in E can be written as a linear combination $c = k_1 b_1 + \dots + k_n b_n$ with coefficients k_i in K . But each k_i because it is in K , is a linear combination

$$k_i = l_{i1} a_1 + \dots + l_{im} a_m$$

with coefficients l_{ij} in F . Substituting,

$$\begin{aligned} c &= (l_{11} a_1 + \dots + l_{1m} a_m) b_1 + \dots + (l_{n1} a_1 + \dots + l_{nm} a_m) b_n \\ &= \sum l_{ij} a_i b_j \end{aligned}$$

and this is a linear combination of the products $a_i b_j$ with coefficient l_{ij} in F .

To prove that $\{a_i b_j\}$ is linearly independent, suppose $\sum l_{ij} a_i b_j = 0$. This can be written as

$$(l_{11} a_1 + \dots + l_{1m} a_m) b_1 + \dots + (l_{n1} a_1 + \dots + l_{nm} a_m) b_n = 0$$

and since $b_1, \dots, b_n$ are independent, $l_{i1} a_1 + \dots + l_{im} a_m = 0$ for each i . But $a_1, \dots, a_m$ are also independent, so every $l_{ij} = 0$. ■

With this result we can now conclude the following:

Theorem 2 Suppose $F \subseteq K \subseteq E$ where E is a finite extension of K and K is a finite extension of F . Then E is a finite extension of F , and

$$[E : F] = [E : K][K : F]$$

This theorem is a powerful tool in our study of fields. It plays a role in field theory analogous to the role of Lagrange's theorem in group theory. See what it says about any two extensions, K and E of a

fixed “base field” F : If K is a subfield of E_9 then the degree of K (over F) divides the degree of E_9 (over F).

If c is algebraic over F , we say that $F(c)$ is obtained by *adjoining* c to F . If c and d are algebraic over F , we may first adjoin c to F , thereby obtaining $F(c)$, and then adjoin d to $F(c)$. The resulting field is denoted $F(c, d)$ and is the smallest field containing F , c and d . [Indeed, any field containing F , c and d must contain $F(c)$, hence also $F(c, d)$.] It does not matter whether we first adjoin c and then d or vice versa.

If $c_1, \dots, c_n$ are algebraic over F , we let $F(c_1, \dots, c_n)$ be the smallest field containing F and $c_1, \dots, c_n$. We call it the field obtained by *adjoining* $c_1, \dots, c_n$ to F . We may form $F(c_1, \dots, c_n)$ step by step, adjoining one c_i at a time, and the order of adjoining the c_i is irrelevant.

An extension $F(c)$ formed by adjoining a single element to F is called a *simple extension* of F . An extension $F(c_1, \dots, c_n)$ formed by adjoining a finite number of elements $c_1, \dots, c_n$ is called an *iterated extension*. It is called “iterated” because it can be formed step by step, one *simple* extension at a time:

$$F \subseteq F(c_1) \subseteq F(c_1, c_2) \subseteq F(c_1, c_2, c_3) \subseteq \dots \subseteq F(c_1, \dots, c_n) \quad (1)$$

If $c_1, \dots, c_n$ are algebraic over F , then by [Theorem 1](#), each extension in Condition (1) is a finite extension. By [Theorem 2](#), $F(c_1, c_2)$ is a finite extension of F ; applying [Theorem 2](#) again, $F(c_1, c_2, c_3)$ is a finite extension of F ; and so on. So finally, if $c_1, \dots, c_n$ are algebraic over F , then $F(c_1, \dots, c_n)$ is a finite extension of F .

Actually, the converse is true too: *every finite extension is an iterated extension*. This is obvious: for if K is a finite extension of F , say an extension of degree n , then K has a basis $\{a_1, \dots, a_n\}$ over F . This means that every element in K is a linear combination of $a_1, \dots, a_n$ with coefficients in F ; but any field containing F and $a_1, \dots, a_n$ obviously contains all the linear combinations of $a_1, \dots, a_n$; hence K is the smallest field containing F and $a_1, \dots, a_n$. That is, $K = F(a_1, \dots, a_n)$.

In fact, if K is a finite extension of F and $K = F(a_1, \dots, a_n)$, then $a_1, \dots, a_n$ have to be *algebraic* over F . This is a consequence of a simple but important little theorem:

Theorem 3 *If K is a finite extension of F , every element of K is algebraic over F*

PROOF: Indeed, suppose K is of degree n over F , and let c be any element of K . Then the set $\{1, c, c^2, \dots, c^n\}$ is linearly dependent, because it has $n + 1$ elements in a vector space K of dimension n . Consequently, there are scalars $a_0, \dots, a_n \in F$, not all zero, such that $a_0 + a_1c + \dots + a_nc^n = 0$. Therefore c is a root of the polynomial $a(x) = a_0 + a_1x + \dots + a_nx^n$ in $F[x]$. ■

Let us sum up: *Every iterated extension $F(c_1, \dots, c_n)$, where $c_1, \dots, c_n$ are algebraic over F , is finite extension of extension of F . Conversely, every finite extension of F is an iterated extension $F(c_1, \dots, c_n)$, where $c_1, \dots, c_n$ are algebraic over F .*

Here is an example of the concepts presented in this chapter. We have already seen that $\mathbb{Q}(\sqrt{2})$ is of degree 2 over $\mathbb{Q}$, and therefore $\mathbb{Q}(\sqrt{2})$ consists of all the numbers $a + b\sqrt{2}$ where $a, b \in \mathbb{Q}$. Observe that $\sqrt{3}$ cannot be in $\mathbb{Q}(\sqrt{2})$; for if it were, we would have $\sqrt{3} = a + b\sqrt{2}$ for rational a and b ; squaring both sides and solving for $\sqrt{2}$ would give us $\sqrt{2} =$ a rational number, which is impossible.

Since $\sqrt{3}$ is not in $\mathbb{Q}(\sqrt{2})$, $\sqrt{3}$ cannot be a root of a polynomial of degree 1 over $\mathbb{Q}(\sqrt{2})$ (such a

polynomial would have to be $x - \sqrt{3}$). But $\sqrt{3}$ is a root of $x^2 - 3$, which is therefore the minimum polynomial of $\sqrt{3}$ over $\mathbb{Q}(\sqrt{2})$. Thus, $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is of degree 2 over $\mathbb{Q}(\sqrt{2})$, and therefore by Theorem 2, $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is of degree 4 over $\mathbb{Q}$.

By the comments preceding Theorem 1, $\{1, \sqrt{2}\}$ is a basis of $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$, and $\{1, \sqrt{3}\}$ is a basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$. Thus, by the lemma of this chapter, $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$. This means that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ consists of all the numbers $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$, for all a, b, c , and d in $\mathbb{Q}$.

For later reference. The technical observation which follows will be needed later.

By the comments immediately preceding Theorem 1, every element of $F(c_1)$ is a linear combination of powers of c_1 , with coefficients in F . That is, every element of $F(c_1)$ is of the form

$$\sum_i k_i c_1^i \quad (2)$$

where the k_i are in F . For the same reason, every element of $F(c_1 c_2)$ is of the form

$$\sum_j l_j c_2^j$$

where the coefficients l_j are in $F(c_1)$. Thus, each coefficient l_j is equal to a sum of the form (2). But then, clearing brackets, it follows that every element of $F(c_1 c_2)$ is of the form

$$\sum_{i,j} k_{ij} c_1^i c_2^j$$

where the coefficients k_{ij} are in F .

If we continue this process, it is easy to see that every element of $F(c_1, c_2, \dots, c_n)$ is a sum of terms of the form

$$k c_1^{i_1} c_2^{i_2} \cdots c_n^{i_n}$$

where the coefficient k of each term is in F .

EXERCISES

A. Examples of Finite Extensions

1 Find a basis for $\mathbb{Q}(i\sqrt{2})$ over $\mathbb{Q}$, and describe the elements of $\mathbb{Q}(i\sqrt{2})$. (See the two examples immediately following Theorem 1.)

2 Show that every element of $\mathbb{R}(2 + 3i)$ can be written as $a + bi$, where $a, b \in \mathbb{R}$. Conclude that $\mathbb{R}(2 + 3i) = \mathbb{C}$.

3 If $a = \sqrt{1 + \sqrt[3]{2}}$, show that $\{1, 2^{1/3}, 2^{2/3}, a, 2^{1/3}a, 2^{2/3}a\}$ is a basis of $\mathbb{Q}(a)$ over $\mathbb{Q}$. Describe the elements of $\mathbb{Q}(a)$.

4 Find a basis of $\mathbb{Q}(\mathbb{Q}(\sqrt{2} + \sqrt[3]{4}))$ over $\mathbb{Q}$, and describe the elements of $\mathbb{Q}(\sqrt{2} + \sqrt[3]{4})$.

5 Find a basis of $\mathbb{Q}(\mathbb{Q}(\sqrt{5}, \sqrt{7}))$ over $\mathbb{Q}$ and describe the elements of $\mathbb{Q}(\mathbb{Q}(\sqrt{5}, \sqrt{7}))$. (See the example at

the end of this chapter.)

6 Find a basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ over $\mathbb{Q}$, and describe the elements of $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$.

7 Name an extension of $\mathbb{Q}$ over which π is algebraic of degree 3.

† B. Further Examples of Finite Extensions

Let F be a field of characteristic $\neq 2$. Let $a \neq b$ be in F .

1 Prove that any field F containing $\sqrt{a} + \sqrt{b}$ also contains $\sqrt{a}$ and $\sqrt{b}$. [HINT: Compute $(\sqrt{a} + \sqrt{b})^2$ and show that $\sqrt{ab} \in F$. Then compute $\sqrt{ab}(\sqrt{a} + \sqrt{b})$, which is also in F] Conclude that $F(\sqrt{a} + \sqrt{b}) = F(\sqrt{a}, \sqrt{b})$.

2 Prove that if $b \neq x^2 a$ for any $x \in F$, then $\sqrt{b} \notin F(\sqrt{a})$. Conclude that $F(\sqrt{a}, \sqrt{b})$ is of degree 4 over F .

3 Show that $x = \sqrt{a} + \sqrt{b}$ satisfies $x^4 - 2(a+b)x^2 + (a-b)^2 = 0$. Show that $x = \sqrt{a+b+2\sqrt{ab}}$ also satisfies this equation. Conclude that

$$F(\sqrt{a+b+2\sqrt{ab}}) = F(\sqrt{a}, \sqrt{b})$$

4 Using parts 1 to 3, find an uncomplicated basis for $\mathbb{Q}(d)$ over $\mathbb{Q}$, where d is a root of $x^4 - 14x^2 + 9$. Then find a basis for $\mathbb{Q}(\sqrt{7+2\sqrt{10}})$ over $\mathbb{Q}$.

C. Finite Extensions of Finite Fields

By the proof of the basic theorem of field extensions, if $p(x)$ is an irreducible polynomial of degree n in $F[x]$, then $F[x]/\langle p(x) \rangle \cong F(c)$ where c is a root of $p(x)$. By [Theorem 1](#) in this chapter, $F(c)$ is of degree n over F . Using the paragraph preceding [Theorem 1](#):

1 Prove that every element of $F(c)$ can be written *uniquely* as $a_0 + a_1 c + \dots + a_{n-1} c^{n-1}$, for some $a_0, \dots, a_{n-1} \in F$.

2 Construct a field of four elements. (It is to be an extension of $\mathbb{Z}_2$.) Describe its elements, and supply its addition and multiplication tables.

3 Construct a field of eight elements. (It is to be an extension of $\mathbb{Z}_2$.)

4 Prove that if F has q elements, and a is algebraic over F of degree n , then $F(a)$ has q^n elements.

5 Prove that for every prime number p , there is an irreducible quadratic in $\mathbb{Z}_p[x]$. Conclude that for every prime p , there is a field with p^2 elements.

D. Degrees of Extensions (Applications of Theorem 2)

Let F be a field, and K a field extension of F . Prove the following:

1 $[K:F] = 1$ iff $K = F$.

2 If $[K:F]$ is a prime number, there is no field properly between F and K (that is, there is no field L such that $F \subsetneq L \subsetneq K$).

3 If $[K:F]$ is a prime, then $K = F(a)$ for every $a \in K - F$.

4 Suppose $a, b \in K$ are algebraic over F with degrees m and n , where m and n are relatively prime. Then:

(a) $F(a,b)$ is degree mn over F .

(b) $F(a) \cap F(b) = F$.

5 If the degree of $F(a)$ over F is a prime, then $F(a) = F(a^n)$ for any n (on the condition that $a^n \notin F$).

6 If an irreducible polynomial $p(x) \in F[x]$ has a root in K , then $\deg p(x) \mid [K:F]$.

E. Short Questions Relating to Degrees of Extensions

Let F be a field.

Prove parts 1–3:

1 The degree of a over F is the same as the degree of $1/a$ over F . It is also the same as the degrees of $a + c$ and ac over F , for any $c \in F$.

2 a is of degree 1 over F iff $a \in F$.

3 If a real number c is a root of an irreducible polynomial of degree > 1 in $\mathbb{Q}[x]$, then c is irrational.

4 Use part 3 and Eisenstein's irreducibility criterion to prove that $\sqrt[m]{m/n}$ (where $m, n \in \mathbb{Z}$) is irrational if there is a prime number which divides m but not n , and whose square does not divide m .

5 Show that part 4 remains true for $\sqrt[q]{m/n}$ where $q > 1$.

6 If a and b are algebraic over F , prove that $F(a, b)$ is a finite extension of F .

† F. Further Properties of Degrees of Extensions

Let F be a field, and K a finite extension of F . Prove each of the following:

1 Any element algebraic over K is algebraic over F , and conversely.

2 If b is algebraic over K , then $[F(b):F] \mid [K(b):F]$.

3 If b is algebraic over K , then $[K(b):K] \leq [F(b):F]$. (HINT: The minimum polynomial of b over F may factor in $K[x]$, and b will then be a root of one of its irreducible factors.)

4 If b is algebraic over K , then $[K(b):F(b)] \leq [K:F]$. [HINT: Note that $F \subseteq K \subseteq K(b)$ and $F \subseteq F(b) \subseteq K(b)$. Relate the degrees of the four extensions involved here, using part 3.]

5 Let $p(x)$ be irreducible in $F[x]$. If $[K:F]$ and $\deg p(x)$ are relatively prime, then $p(x)$ is irreducible in $K[x]$.

† G. Fields of Algebraic Elements: Algebraic Numbers

Let $F \subseteq K$ and $a, b \in K$. We have seen on page 295 that if a and b are algebraic over F , then $F(a, b)$ is a finite extension of F .

Use the above to prove parts 1 and 2.

1 If a and b are algebraic over F , then $a + b$, $a - b$, ab , and a/b are algebraic over F . (In the last case, assume $b \neq 0$.)

2 The set $\{x \in K : x \text{ is algebraic over } F\}$ is a subfield of K , containing F .

Any complex number which is algebraic over $\mathbb{Q}$ is called an *algebraic number*. By part 2, the set of all the algebraic numbers is a field, which we shall designate by $\mathbb{A}$.

Let $a(x) = a_0 + a_1x + \dots + a_nx^n$ be in $\mathbb{A}[x]$, and let c be any root of $a(x)$. We will prove that $c \in \mathbb{A}$. To begin with, all the coefficients of $a(x)$ are in $\mathbb{Q}(a_0, a_1, \dots, a_n)$.

3 Prove: $\mathbb{Q}(a_0, a_1, \dots, a_n)$ is a finite extension of $\mathbb{Q}$.

Let $\mathbb{Q}(a_0, \dots, a_n) = \mathbb{Q}_1$ Since $a(x) \in \mathbb{Q}_1[x]$, c is algebraic over $\mathbb{Q}_1$ Prove parts 4 and 5:

4 $\mathbb{Q}_1(c)$ is a finite extension of $\mathbb{Q}_1$ hence a finite extension of $\mathbb{Q}$. (Why?)

5 $c \in \mathbb{A}$.

Conclusion: The roots of any polynomial whose coefficients are algebraic numbers are themselves algebraic numbers.

A field F is called *algebraically closed* if the roots of every polynomial in $F[x]$ are in F . We have thus proved that $\mathbb{A}$ is algebraically closed.

RULER AND COMPASS

The ancient Greek geometers considered the circle and straight line to be the most basic of all geometric figures, other figures being merely variants and combinations of these basic ones. To understand this view we must remember that construction played a very important role in Greek geometry: when a figure was defined, a method was also given for constructing it. Certainly the circle and the straight line are the easiest figures to construct, for they require only the most rudimentary of all geometric instruments: the ruler and the compass. Furthermore, the ruler, in this case, is a simple, unmarked straightedge.

Rudimentary as these instruments may be, they can be used to carry out a surprising variety of geometric constructions. Lines can be divided into any number of equal segments, and any angle can be bisected. From any polygon it is possible to construct a square having the same area, or twice or three times the area. With amazing ingenuity, Greek geometers devised ways to cleverly use the ruler and compass, unaided by any other instrument, to perform all kinds of intricate and beautiful constructions. They were so successful that it was hard to believe they were unable to perform three little tasks which, at first sight, appear to be very simple: *doubling the cube*, *trisecting any angle*, and *squaring the circle*. The first task demands that a cube be constructed having twice the volume of a given cube. The second asks that any angle be divided into three equal parts. The third requires the construction of a square whose area is equal to that of a given circle. Remember, only a ruler and compass are to be used!

Mathematicians, in Greek antiquity and throughout the Renaissance, devoted a great deal of attention to these problems, and came up with many brilliant ideas. But they never found ways of performing the above three constructions. This is not surprising, for these constructions are impossible! Of course, the Greeks had no way of knowing that fact, for the mathematical machinery needed to prove that these constructions are impossible—in fact, the very notion that one could prove a construction to be impossible—was still two millennia away.

The final resolution of these problems, by proving that the required constructions are impossible, came from a most unlikely source: it was a by-product of the arcane study of field extensions, in the upper reaches of modern algebra.

To understand how all this works, we will see how the process of ruler-and-compass constructions can be placed in the framework of field theory. Clearly, we will be making use of analytic geometry.

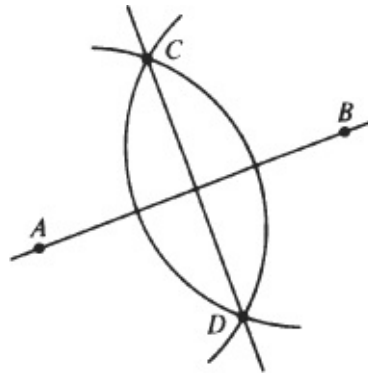
If $\mathcal{A}$ is any set of points in the plane, consider operations of the following two kinds:

1. *Ruler operation*: Through any two points in $\mathcal{A}$, draw a straight line.

2. *Compass operation:* Given three points A , B , and C in $\mathcal{A}$, draw a circle with center C and radius equal in length to the segment AB .

The points of intersection of any two of these figures (line-line, line-circle, or circle-circle) are said to be *constructible in one step* from $\mathcal{A}$. A point P is called *constructible* from $\mathcal{A}$ if there are points $P_1, P_2, \dots, P_n = P$ such that P_1 is constructible in one step from $\mathcal{A}$, P_2 is constructible in one step from $\mathcal{A} \cup \{P_1\}$, and so on, so that P_i is constructible in one step from $\mathcal{A} \cup \{P_1, \dots, P_{i-1}\}$.

As a simple example, let us see that the midpoint of a line segment AB is constructible from the two points A and B in the above sense. Well, given A and B , first draw the line AB . Then, draw the circle with center A and radius $\overline{AB}$ and the circle with center B and radius $\overline{AB}$; let C and D be the points of intersection of these circles. C and D are constructible in one step from $\{A, B\}$. Finally, draw the line through C and D ; the intersection of this line with AB is the required midpoint. It is constructible from $\{A, B\}$.



As this example shows, the notion of constructible points is the correct formalization of the intuitive idea of ruler-and-compass constructions.

We call a point in the plane *constructible* if it is constructible from $\mathbb{Q} \times \mathbb{Q}$, that is, from the set of all points in the plane with rational coefficients.

How does field theory fit into this scheme? Obviously by associating with every point its coordinates. More exactly, with every constructible point P we associate a certain field extension of $\mathbb{Q}$, obtained as follows:

Suppose P has coordinates (a, b) and is constructed from $\mathbb{Q} \times \mathbb{Q}$ in one step. We associate with P the field $\mathbb{Q}(a, b)$, obtained by adjoining to $\mathbb{Q}$ the coordinates of P . More generally, suppose P is constructible from $\mathbb{Q} \times \mathbb{Q}$ in n steps: there are then n points $P_1, P_2, \dots, P_n = P$ such that each P_i is constructible in one step from $\mathbb{Q} \times \mathbb{Q} \cup \{P_1, \dots, P_{i-1}\}$. Let the coordinates of $P_1, \dots, P_n$ be $(a_1, b_1), \dots, (a_n, b_n)$, respectively. With the points $P_1, \dots, P_n$ we associate fields $K_1, \dots, K_n$ where $K_1 = \mathbb{Q}(a_1, b_1)$, and for each $i > 1$,

$$K_i = K_{i-1}(a_i, b_i)$$

Thus, $K_1 = \mathbb{Q}(a_1, b_1)$, $K_2 = K_1(a_2, b_2)$, and so on: beginning with $\mathbb{Q}$, we adjoin first the coordinates of P_1 , then the coordinates of P_2 , and so on successively, yielding the sequence of extensions

$$\mathbb{Q} \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n = K$$

We call K the *field extension associated with the point P* .

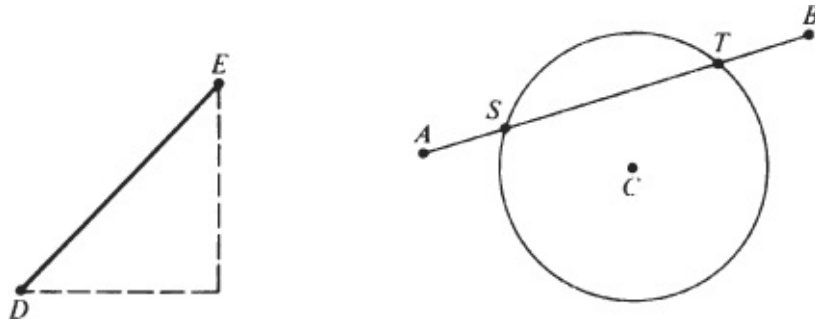
Everything we will have to say in the sequel follows easily from the next lemma.

Lemma *If $K_1, \dots, K_n$ are as defined previously, then $[K_i : K_{i-1}] = 1, 2, \text{ or } 4$.*

PROOF: Remember that K_{i-1} already contains the coordinates of $P_1, \dots, P_{i-1}$, and K_i is obtained by adjoining to K_{i-1} the coordinates x_i, y_i of P_i . But P_i is constructible in one step from $\mathbb{Q} \times \mathbb{Q} \cup \{P_1, \dots, P_{i-1}\}$, so we must consider three cases, corresponding to the three kinds of intersection which may produce P_i , namely: line intersects line, line intersects circle, and circle intersects circle.

Line intersects line: Suppose one line passes through the points (a_1, a_2) and (b_1, b_2) , and the other line passes through (c_1, c_2) and (d_1, d_2) . We may write equations for these lines in terms of the constants $a_1, a_2, b_1, b_2, c_1, c_2$ and d_1, d_2 (all of which are in K_{i-1}), and then solve these equations simultaneously to give the coordinates x, y of the point of intersection. Clearly, these values of x and y are expressed in terms of $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$, hence are still in K_{i-1} . Thus, $K_i = K_{i-1}$.

Line intersects circle: Consider the line AB and the circle with center C and radius equal to the distance $k = \overline{DE}$. Let A, B, C have coordinates $(a_1, a_2), (b_1, b_2)$, and (c_1, c_2) , respectively. By hypothesis, K_{i-1} contains the numbers $a_1, a_2, b_1, b_2, c_1, c_2$, as well as $k^2 = \text{the square of the distance } \overline{DE}$. (To understand the last assertion, remember that K_{i-1} contains the coordinates of D and E ; see the figure and use the Pythagorean theorem.)



Now, the line AB has equation

$$\frac{y - b_2}{x - b_1} = \frac{b_2 - a_2}{b_1 - a_1} \quad (1)$$

and the circle has equation

$$(x - c_1)^2 + (y - c_2)^2 = k^2 \quad (2)$$

Solving for x in (1) and substituting into (2) gives

$$(x - c_1)^2 + \frac{b_2 - a_2}{b_1 - a_1} (x - b_1) - b_2^2 - c_2^2 = k^2$$

This is obviously a quadratic equation, and its roots are the x coordinates of S and T . Thus, the x coordinates of both points of intersection are roots of a quadratic polynomial with coefficients in K_{i-1} . The same is true of the y coordinates. Thus, if $K_i = K_{i-1}(x_i, y_i)$ where (x_i, y_i) is one of the points of intersection, then

$$[K_{i-1}(x_i, y_i) : K_{i-1}] = [K_{i-1}(x_i, y_i) : K_{i-1}(x_i)][K_{i-1}(x_i) : K_{i-1}]$$

$$= 2 \times 2 = 4$$

{This assumes that $x_i, y_i \notin K_{i-1}$. If either x_i or y_i , or both are already in K_{i-1} , then $[K_{i-1}(x_i, y_i) : K_{i-1}] = 1$ or 2 .}

Circle intersects circle: Suppose the two circles have equations

$$x^2 + y^2 + ax + by + c = 0 \quad (3)$$

and

$$x^2 + y^2 + dx + ey + f = 0 \quad (4)$$

Then both points of intersection satisfy

$$(a - d)x + (b - e)y + (c - f) = 0 \quad (5)$$

obtained simply by subtracting (4) from (3). Thus, x and y may be found by solving (4) and (5) simultaneously, which is exactly the preceding case. ■

We are now in a position to prove the main result of this chapter:

Theorem 1: Basic theorem on constructible points *If the point with coordinates (a, b) is constructible, then the degree of $\mathbb{Q}(a)$ over $\mathbb{Q}$ is a power of 2, and likewise for the degree of $\mathbb{Q}(b)$ over $\mathbb{Q}$.*

PROOF: Let P be a constructible point; by definition, there are points $P_1, \dots, P_n$ with coordinates $(a_1, b_1), \dots, (a_n, b_n)$ such that each P_i is constructible in one step from $\mathbb{Q} \times \mathbb{Q} \cup \{P_1, \dots, P_{i-1}\}$, and $P_n = P$. Let the fields associated with $P_1, \dots, P_n$ be $K_1, \dots, K_n$. Then

$$[K_n : \mathbb{Q}] = [K_n : K_{n-1}] [K_{n-1} : K_{n-2}] \cdots [K_1 : \mathbb{Q}]$$

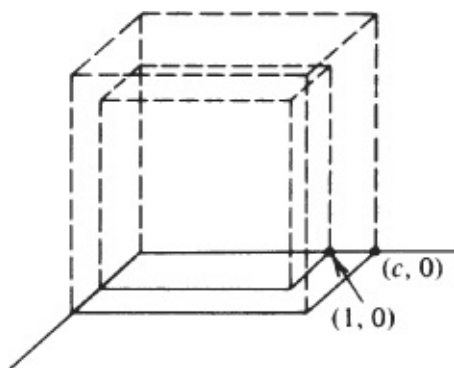
and by the preceding lemma this is a power of 2, say 2^m . But

$$[K_n : \mathbb{Q}] = [K_n : \mathbb{Q}(a)][\mathbb{Q}(a) : \mathbb{Q}]$$

hence $[\mathbb{Q}(a) : \mathbb{Q}]$ is a factor of 2^m , hence also a power of 2. ■

We will now use this theorem to prove that ruler-and-compass constructions cannot possibly exist for the three classical problems described in the opening to this chapter.

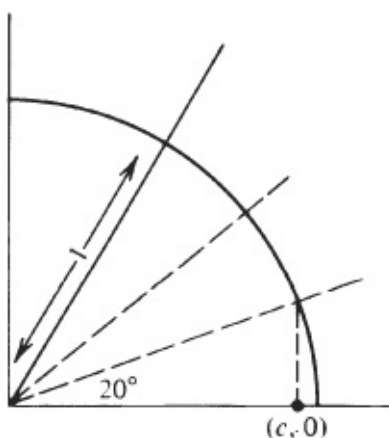
Theorem 2 *“Doubling the cube” is impossible by ruler and compass.*



PROOF: Let us place the cube on a coordinate system so that one edge of the cube coincides with the unit interval on the x axis. That is, its endpoints are $(0,0)$ and $(1,0)$. If we were able to double the cube by ruler and compass, this means we could construct a point $(c, 0)$ such that $c^3 = 2$. However, by [Theorem 1](#), $[\mathbb{Q}(c) : \mathbb{Q}]$ would have to be a power of 2, whereas in fact it is obviously 3. This contradiction proves that it is impossible to double the cube using only a ruler and compass. ■

Theorem 3 “Trisecting the angle” by ruler and compass is impossible. That is, there exist angles which cannot be trisected using a ruler and compass.

PROOF: We will show specifically that an angle of 60° cannot be trisected. If we *could* trisect an angle of 60° , we would be able to construct a point $(c, 0)$ (see figure) where $c = \cos 20^\circ$; hence certainly we could construct $(b, 0)$ where $b = 2 \cos 20^\circ$.



But from elementary trigonometry

$$\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta$$

hence

$$\underbrace{\cos 60^\circ}_{\frac{1}{2}} = 4 \cos^3 20^\circ - 3 \cos 20^\circ$$

Thus, $b = 2 \cos 20^\circ$ satisfies $b^3 - 3b - 1 = 0$. The polynomial

$$p(x) = x^3 - 3x - 1$$

is irreducible over $\mathbb{Q}$ because $p(x+1) = x^3 + 3x^2 - 3$ is irreducible by Eisenstein’s criterion. It follows that $\mathbb{Q}(b)$ has degree 3 over $\mathbb{Q}$, contradicting the requirement (in [Theorem 1](#)) that this degree has to be a power of 2. ■

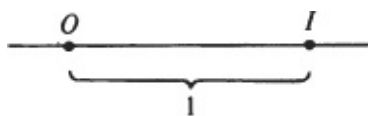
Theorem 4 “Squaring the circle” by ruler and compass is impossible.

PROOF. If we were able to square the circle by ruler and compass, it would be possible to construct the point $(0, \sqrt{\pi})$; hence by Theorem 1, $[\mathbb{Q}(\sqrt{\pi}) : \mathbb{Q}]$ would be a power of 2. But it is well known that π is transcendental over $\mathbb{Q}$. By Theorem 3 of Chapter 29, the square of an algebraic element is algebraic; hence $\sqrt{\pi}$ is transcendental. It follows that $\mathbb{Q}(\sqrt{\pi})$ is not even a finite extension of $\mathbb{Q}$, much less an extension of some degree 2^m as required. ■

EXERCISES

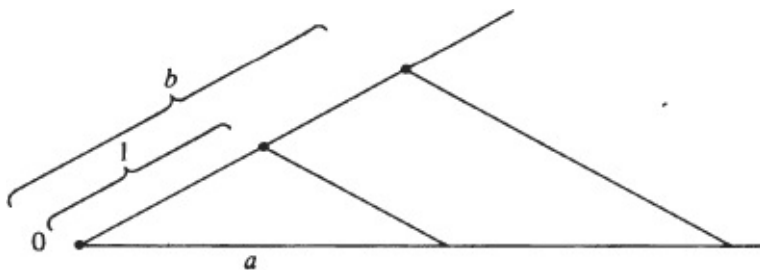
† A. Constructible Numbers

If O and I are any two points in the plane, consider a coordinate system such that

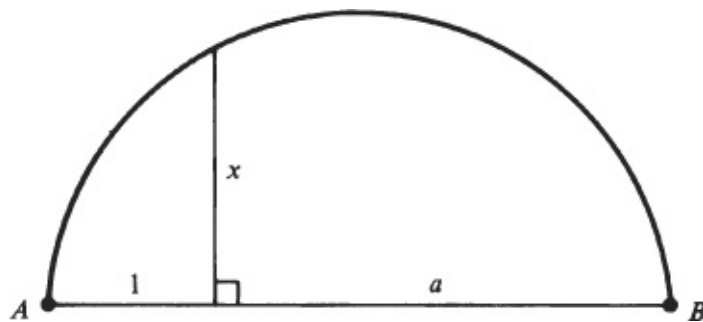


the interval OI coincides with the unit interval on the x axis. Let $\mathbb{D}$ be the set of real numbers such that $a \in \mathbb{D}$ iff the point $(a, 0)$ is constructible from $\{O, I\}$. Prove the following:

- 1 If $a, b \in \mathbb{D}$, then $a + b \in \mathbb{D}$ and $a - b \in \mathbb{D}$.
- 2 If $a, b \in \mathbb{D}$, then $ab \in \mathbb{D}$. (HINT: Use similar triangles. See the accompanying figure.)



- 3 If $a, b \in \mathbb{D}$, then $a/b \in \mathbb{D}$. (Use the same figure as in part 2.)
- 4 If $a > 0$ and $a \in \mathbb{D}$, then $\sqrt{a} \in \mathbb{D}$. (HINT: In the accompanying figure, AB is the diameter of a circle. Use an elementary property of chords of a circle to show that $x = \sqrt{a}$.)



It follows from parts 1 to 4 that $\mathbb{D}$ is a field, closed with respect to taking square roots of positive numbers. $\mathbb{D}$ is called the *field of constructible numbers*.

5 $\mathbb{Q} \subseteq \mathbb{D}$.

6 If a is a real root of any quadratic polynomial with coefficients in $\mathbb{D}$, then $a \in \mathbb{D}$. (HINT: Complete the square and use part 4.)

† B. Constructible Points and Constructible Numbers

Prove each of the following:

1 Let $\mathcal{A}$ be any set of points in the plane; (a, b) is constructible from $\mathcal{A}$ iff $(a, 0)$ and $(0, b)$ are constructible from $\mathcal{A}$.

2 If a point P is constructible from $\{O, I\}$ [that is, from $(0, 0)$ and $(1, 0)$], then P is constructible from $\mathbb{Q} \times \mathbb{Q}$.

3 Every point in $\mathbb{Q} \times \mathbb{Q}$ is constructible from $\{O, I\}$. (Use [Exercise A5](#) and the definition of $\mathbb{D}$.)

4 If a point P is constructible from $\mathbb{Q} \times \mathbb{Q}$, it is constructible from $\{O, I\}$.

By combining parts 2 and 4, we get the following important fact: Any point P is constructible from $\mathbb{Q} \times \mathbb{Q}$ iff P is constructible from $\{O, I\}$. Thus, we may define a point to be *constructible* iff it is constructible from $\{O, I\}$.

5 A point P is constructible iff both its coordinates are constructible numbers.

† C. Constructible Angles

An angle α is called *constructible* iff there exist constructible points A, B , and C such that $\angle ABC = \alpha$. Prove the following:

1 The angle α is constructible iff $\sin \alpha$ and $\cos \alpha$ are constructible numbers.

2 $\cos \alpha \in \mathbb{D}$ iff $\sin \alpha \in \mathbb{D}$.

3 If $\cos \alpha, \cos \beta \in \mathbb{D}$, then $\cos(\alpha + \beta), \cos(\alpha - \beta) \in \mathbb{D}$.

4 $\cos(2\alpha) \in \mathbb{D}$ iff $\cos \alpha \in \mathbb{D}$.

5 If α and β are constructible angles, so are $\alpha + \beta, \alpha - \beta, \frac{1}{2}\alpha$, and $n\alpha$ for any positive integer n .

6 The following angles are constructible: $30^\circ, 75^\circ, 22\frac{1}{2}^\circ$.

7 The following angles are *not* constructible: $20^\circ, 40^\circ, 140^\circ$. (HINT: Use the proof of [Theorem 3](#).)

D. Constructible Polygons

A polygon is called *constructible* iff its vertices are constructible points. Prove the following:

1 The regular n -gon is constructible iff the angle $2\pi/n$ is constructible.

2 The regular hexagon is constructible.

3 The regular polygon of nine sides is *not* constructible.

† E. A Constructible Polygon

We will show that $2\pi/5$ is a constructible angle, and it will follow that the regular pentagon is constructible.

1 If $r = \cos k + i \sin k$ is a complex number, prove that $1/r = \cos k - i \sin k$. Conclude that $r + 1/r = 2 \cos k$.

By de Moivre's theorem,

$$\omega = \cos \frac{2\pi}{5} + i \sin \frac{2\pi}{5}$$

is a complex fifth root of unity. Since

$$x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1)$$

ω is a root of $p(x) = x^4 + x^3 + x^2 + x + 1$.

2 Prove that $\omega^2 + \omega + 1 + \omega^{-1} + \omega^{-2} = 0$.

3 Prove that

$$4 \cos^2 \frac{2\pi}{5} + 2 \cos \frac{2\pi}{5} - 1 = 0$$

(HINT: Use parts 1 and 2.) Conclude that $\cos(2\pi/5)$ is a root of the quadratic $4x^2 - 2x - 1$.

4 Use part 3 and A6 to prove that $\cos(2\pi/5)$ is a constructible number.

5 Prove that $2\pi/5$ is a constructible angle.

6 Prove that the regular pentagon is constructible.

† F. A Nonconstructible Polygon

By de Moivre's theorem,

$$\omega = \cos \frac{2\pi}{7} + i \sin \frac{2\pi}{7}$$

is a complex seventh root of unity. Since

$$x^7 - 1 = (x - 1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1)$$

ω is a root of $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$.

1 Prove that $\omega^3 + \omega^2 + \omega + 1 + \omega^{-1} + \omega^{-2} + \omega^{-3} = 0$.

2 Prove that

$$8 \cos^3 \frac{2\pi}{7} + 4 \cos^2 \frac{2\pi}{7} - 4 \cos \frac{2\pi}{7} - 1 = 0$$

(Use part 1 and [Exercise E1](#).) Conclude that $\cos(2\pi/7)$ is a root of $8x^3 + 4x^2 - 4x - 1$.

3 Prove that $8x^3 + 4x^2 - 4x - 1$ has no rational roots. Conclude that it is irreducible over $\mathbb{Q}$.

4 Conclude from part 3 that $\cos(2\pi/7)$ is not a constructible number.

5 Prove that $2\pi/7$ is not a constructible angle.

6 Prove that the regular polygon of seven sides is not constructible.

G. Further Properties of Constructible Numbers and Figures

Prove each of the following:

1 If the number a is a root of an irreducible polynomial $p(x) \in \mathbb{Q}[x]$ whose degree is not a power of 2, then a is not a constructible number.

2 Any constructible number can be obtained from rational numbers by repeated addition, subtraction,

multiplication, division, and taking square roots of positive numbers.

3 $\mathbb{D}$ is the smallest field extension of $\mathbb{Q}$ closed with respect to square roots of positive numbers (that is, any field extension of $\mathbb{Q}$ closed with respect to square roots contains $\mathbb{D}$). (Use part 2 and [Exercise A](#).)

4 All the roots of the polynomial $x^4 - 3x^2 + 1$ are constructible numbers.

A line is called constructible if it passes through two constructible points. A circle is called constructible if its center and radius are constructible.

5 The line $ax + by + c = 0$ is constructible if $a, b, c \in \mathbb{D}$.

6 The circle $x^2 + y^2 + ax + by + c = 0$ is constructible if $a, b, c \in \mathbb{D}$.