

ALGEBRA

SECOND EDITION

SAUNDERS MACLANE

Max Mason Distinguished Service Professor of Mathematics
The University of Chicago

GARRETT BIRKHOFF

Putnam Professor of Pure and Applied Mathematics
Harvard University

MACMILLAN PUBLISHING CO., INC.

New York

COLLIER MACMILLAN PUBLISHERS

Copyright © 1979, Macmillan Publishing Co., Inc.

Printed in the United States of America

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

Earlier edition Copyright © 1967 by
Macmillan Publishing Co., Inc.

Macmillan Publishing Co., Inc.
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Ltd.

Library of Congress Cataloging in Publication Data

Mac Lane, Saunders, (date)
Algebra.

Includes bibliographies and index.

1. Algebra, Abstract. I. Birkhoff, Garrett, (date)
joint author, II. Title.

QA162.M33 1979

512'.02

78-6946

ISBN 0-02-374310-7

Printing: 1 2 3 4 5 6 7 8

Year: 9 0 1 2 3 4 5

Contents

LIST OF SYMBOLS

xiii

CHAPTER

I	<i>Sets, Functions, and Integers</i>	1
	1. Sets • 1	
	2. Functions • 4	
	3. Relations and Binary Operations • 10	
	4. The Natural Numbers • 15	
	5. Addition and Multiplication • 18	
	6. Inequalities • 20	
	7. The Integers • 23	
	8. The Integers Modulo n • 28	
	9. Equivalence Relations and Quotient Sets • 33	
	10. Morphisms • 37	
	11. Semigroups and Monoids • 39	
II	<i>Groups</i>	43
	1. Groups and Symmetry • 43	
	2. Rules of Calculation • 47	
	3. Cyclic Groups • 51	
	4. Subgroups • 56	
	5. Defining Relations • 59	
	6. Symmetric and Alternating Groups • 63	
	7. Transformation Groups • 68	
	8. Cosets • 72	
	9. Kernel and Image • 75	
	10. Quotient Groups • 79	
III	<i>Rings</i>	85
	1. Axioms for Rings • 85	
	2. Constructions for Rings • 90	
	3. Quotient Rings • 95	
	4. Integral Domains and Fields • 99	
	5. The Field of Quotients • 101	
	6. Polynomials • 104	

7. Polynomials as Functions • 109
8. The Division Algorithm • 111
9. Principal Ideal Domains • 115
10. Unique Factorization • 116
11. Prime Fields • 120
12. The Euclidean Algorithm • 122
13. Commutative Quotient Rings • 124

IV *Universal Constructions* 129

1. Examples of Universals • 129
2. Functors • 131
3. Universal Elements • 134
4. Polynomials in Several Variables • 137
5. Categories • 141
6. Posets and Lattices • 143
7. Contravariance and Duality • 146
8. The Category of Sets • 153
9. The Category of Finite Sets • 156

V *Modules* 160

1. Sample Modules • 160
2. Linear Transformations • 163
3. Submodules • 167
4. Quotient Modules • 171
5. Free Modules • 173
6. Biproducts • 178
7. Dual Modules • 185

VI *Vector Spaces* 193

1. Bases and Coordinates • 194
2. Dimension • 199
3. Constructions for Bases • 202
4. Dually Paired Vector Spaces • 207
5. Elementary Operations • 212
6. Systems of Linear Equations • 219

VII *Matrices* 223

1. Matrices and Free Modules • 224
2. Matrices and Biproducts • 232
3. The Matrix of a Map • 236
4. The Matrix of a Composite • 240
5. Ranks of Matrices • 244

- 6. Invertible Matrices • 246
- 7. Change of Bases • 251
- 8. Eigenvectors and Eigenvalues • 257

VIII *Special Fields* 261

- 1. Ordered Domains • 261
- 2. The Ordered Field \mathbb{Q} • 265
- 3. Polynomial Equations • 267
- 4. Convergence in Ordered Fields • 269
- 5. The Real Field \mathbb{R} • 271
- 6. Polynomials over \mathbb{F} • 274
- 7. The Complex Plane • 276
- 8. The Quaternions • 281
- 9. Extended Formal Power Series • 284
- 10. Valuations and p -adic Numbers • 286

IX *Determinants and Tensor Products* 293

- 1. Multilinear and Alternating Functions • 293
- 2. Determinants of Matrices • 296
- 3. Cofactors and Cramer's Rule • 301
- 4. Determinants of Maps • 305
- 5. The Characteristic Polynomial • 309
- 6. The Minimal Polynomial • 312
- 7. Universal Bilinear Functions • 318
- 8. Tensor Products • 319
- 9. Exact Sequences • 326
- 10. Identities on Tensor Products • 329
- 11. Change of Rings • 331
- 12. Algebras • 334

X *Bilinear and Quadratic Forms* 338

- 1. Bilinear Forms • 338
- 2. Symmetric Matrices • 341
- 3. Quadratic Forms • 343
- 4. Real Quadratic Forms • 347
- 5. Inner Products • 351
- 6. Orthonormal Bases • 355
- 7. Orthogonal Matrices • 360
- 8. The Principal Axis Theorem • 364
- 9. Unitary Spaces • 369
- 10. Normal Matrices • 374

XI	<i>Similar Matrices and Finite Abelian Groups</i>	378
	1. Noetherian Modules • 378	
	2. Cyclic Modules • 381	
	3. Torsion Modules • 383	
	4. The Rational Canonical Form for Matrices • 388	
	5. Primary Modules • 392	
	6. Free Modules • 397	
	7. Equivalence of Matrices • 400	
	8. The Calculation of Invariant Factors • 404	
XII	<i>Structure of Groups</i>	409
	1. Isomorphism Theorems • 409	
	2. Group Extensions • 413	
	3. Characteristic Subgroups • 417	
	4. Conjugate Classes • 419	
	5. The Sylow Theorems • 422	
	6. Nilpotent Groups • 426	
	7. Solvable Groups • 428	
	8. The Jordan–Hölder Theorem • 430	
	9. Simplicity of A_n • 433	
XIII	<i>Galois Theory</i>	436
	1. Quadratic and Cubic Equations • 436	
	2. Algebraic and Transcendental Elements • 439	
	3. Degrees • 442	
	4. Ruler and Compass • 445	
	5. Splitting Fields • 446	
	6. Galois Groups of Polynomials • 450	
	7. Separable Polynomials • 453	
	8. Finite Fields • 456	
	9. Normal Extensions • 458	
	10. The Fundamental Theorem • 462	
	11. The Solution of Equations by Radicals • 465	
XIV	<i>Lattices</i>	470
	1. Posets: Duality Principle • 470	
	2. Lattice Identities • 473	
	3. Sublattices and Products of Lattices • 476	
	4. Modular Lattices • 478	
	5. Jordan–Hölder–Dedekind Theorem • 480	
	6. Distributive Lattices • 483	

7. Rings of Sets	• 485
8. Boolean Algebras	• 487
9. Free Boolean Algebras	• 491
XV	<i>Categories and Adjoint Functors</i>
	495
1. Categories	• 495
2. Functors	• 501
3. Contravariant Functors	• 504
4. Natural Transformations	• 506
5. Representable Functors and Universal Elements	• 511
6. Adjoint Functors	• 517
XVI	<i>Multilinear Algebra</i>
	522
1. Iterated Tensor Products	• 522
2. Spaces of Tensors	• 524
3. Graded Modules	• 530
4. Graded Algebras	• 533
5. The Graded Tensor Algebra	• 539
6. The Exterior Algebra of a Module	• 543
7. Determinants by Exterior Algebra	• 547
8. Subspaces by Exterior Algebra	• 552
9. Duality in Exterior Algebra	• 555
10. Alternating Forms and Skew-Symmetric Tensors	• 558
	<i>Bibliography</i>
	561
	<i>Index</i>
	565

CHAPTER I

Sets, Functions, and Integers

ALGEBRA starts as the art of manipulating sums, products, and powers of numbers. The rules for these manipulations hold for all numbers, so the manipulations may be carried out with letters standing for the numbers. It then appears that the same rules hold for various different sorts of numbers, rational, real, or complex, and that the rules for multiplication even apply to things such as transformations which are not numbers at all. An algebraic system, as we will study it, is thus a set of elements of any sort on which functions such as addition and multiplication operate, provided only that these operations satisfy certain basic rules. The rules for multiplication and inverse are the axioms for a "group", those for addition, subtraction, and multiplication are the axioms for a "ring", and the functions mapping one system to another are the "morphisms". This chapter starts with the necessary ideas about sets, functions, and relations. Then the natural numbers are used to construct the integers and the integers modulo n , with their addition and multiplication. This serves as an introduction to the notion of a morphism from one algebraic system to another.

Many developments in algebra depend vitally upon defining the right concept. When our presentation reaches any definition, the term being defined is put in italics, as *group*, *ring*, *field*, and so on. However, terms little used in the sequel as well as terminology alternative to that selected here are put in quotation marks; thus "range" stands for *codomain* and "onto" for *surjective* (see §2 below).

A reference such as Theorem 3 is to Theorem 3 of the current chapter, while Theorem II.3 is to Theorem 3 of Chapter II. In like manner, Corollary IV.5.2 refers to Corollary 2 of Theorem 5 of Chapter IV, and Equation (VI.11) to Equation (11) of Chapter VI. Within each Chapter, Theorems and Propositions are numbered in a single series. More difficult exercises and sections which may be omitted on first reading are denoted by an asterisk, *.

1. Sets

Intuitively, a "set" is any collection of elements, and a "function" is any rule which assigns to each element of one set a corresponding element of a second set.

Examples of sets abound: The set of all lines in the plane, the set Q of all rational numbers, the set C of all complex numbers, the set Z of all integers (positive, negative, or zero). Sets with only a finite number of different elements may be described by listing all their elements, often indicated by writing these elements between braces. Thus the set of all even integers between 0 and 8, inclusive, may be exhibited as $\{0, 2, 4, 6, 8\}$, while the set of all positive divisors of 6 is the set $\{1, 2, 3, 6\}$. The order in which the elements of a set are listed is irrelevant: $\{1, 3, 6, 2\} = \{1, 2, 3, 6\}$.

More formally, " $x \in S$ " stands for " x is an *element* of the set S " or equivalently, " x is a *member* of the set S " or " x *belongs to* S ". Also, $x \notin S$ means that x is *not* an element of S . Since a set is completely determined by giving its elements, two sets S and T are *equal* if and only if they have the same elements; in symbols:

$$S = T \Leftrightarrow \text{For all } x, x \in S \text{ if and only if } x \in T. \quad (1)$$

(Here the two-pointed double arrow " \Leftrightarrow " stands for "if and only if".) Also, S is a *subset* of T (or, is *included* in T) when every element of S is an element of T , so that, if $x \in S$, then $x \in T$; in symbols:

$$S \subset T \Leftrightarrow \text{For all } x, x \in S \Rightarrow x \in T.$$

(Here, on the right, the one-pointed double arrow " \Rightarrow " stands for "implies".) By this definition, $S \subset T$ and $T \subset U$ imply $S \subset U$, while the equality of sets, as defined above, may be rewritten as

$$S = T \Leftrightarrow S \subset T \text{ and } T \subset S.$$

A set S is *empty* if it has no elements. By the equality rule (1), any two empty sets are equal. Hence, we speak of *the* empty set, written \emptyset . It is also called the *null set* or the *void set*; it is a subset of every set. Also, S is a *proper subset* of a set U when $S \subset U$ but $S \neq \emptyset$ and $S \neq U$.

A particular subset of a given set U is often described as the set of all those elements x in U which have a specified property. Thus the subset of those complex numbers z such that $z^2 = -1$ is written $\{z | z \in C \text{ and } z^2 = -1\}$, while the formulas

$$E = \{x | x \in Z \text{ and } x = 2y \text{ for some } y \in Z\}, \quad N = \{x | x \in Z \text{ and } x > 0\}$$

describe the set E of all even integers and the set N of all nonnegative integers, respectively. Different properties may describe the same subset; thus

$$\{n | n \in Z \text{ and } 0 < n < 1\} \quad \text{and} \quad \{n | n \in Z \text{ and } n^2 = -1\}$$

both describe the empty set \emptyset .

Next we consider the operations of intersection and union on sets. If R and S are given sets, their *intersection* $R \cap S$ is the set of all elements

common to R and S :

$$R \cap S = \{x | x \in R \text{ and } x \in S\},$$

while their *union* $R \cup S$ is the set of all elements which belong either to R or to S (or to both):

$$R \cup S = \{x | x \in R \text{ or } x \in S\}.$$

These definitions may be stated thus:

$$x \in (R \cap S) \Leftrightarrow x \in R \text{ and } x \in S,$$

$$x \in (R \cup S) \Leftrightarrow x \in R \text{ or } x \in S.$$

This display correlates the operations of intersection and union with the logical connectives "and" and "or". The corresponding correlate of "not" is the operation of "complement": If S is a subset of U , the *complement* S' of S in U is the set of all those elements of U which do not belong to S :

$$S' = \{x | x \in U \text{ and } x \notin S\}.$$

For example, for the sets E and N above, $E \cap N$ is the set of even nonnegative integers, $E \cup N$ the set of all integers except the negative odd ones, while the complement E' of E in \mathbb{Z} is the set of all odd integers.

The operations of intersection, union, and complement satisfy various "identities", valid for arbitrary sets. A sample such identity is

$$R \cap (S \cup T) = (R \cap S) \cup (R \cap T), \quad (2)$$

valid for any three sets R , S , and T . (This equation states that the operation "intersection" is distributive over the operation "union".) To prove this statement, consider any element x . By the definitions of \cap and \cup above,

$$\begin{aligned} x \in [R \cap (S \cup T)] &\Leftrightarrow x \in R \text{ and } x \in S \cup T \\ &\Leftrightarrow x \in R \text{ and } (x \in S \text{ or } x \in T). \end{aligned}$$

For similar reasons,

$$x \in [(R \cap S) \cup (R \cap T)] \Leftrightarrow (x \in R \text{ and } x \in S) \text{ or } (x \in R \text{ and } x \in T).$$

Now, in view of familiar properties of "and" and "or", the two different statements made about x at the right of the two displays above are logically equivalent. Hence, the two sets in question have the same elements and therefore are equal. In other words, this proof reduces property (2) of intersection and union to an exactly corresponding property of the logical connectives "and" and "or".

A similar argument gives another distributive law,

$$R \cup (S \cap T) = (R \cup S) \cap (R \cup T). \quad (3)$$

Other algebraic properties of intersection, union, and complement will be considered in the exercises in §3 below.

Two sets, R and S , are called *disjoint* when $R \cap S = \emptyset$.

Given a set U , the set $P(U)$ of all subsets S of U is called the *power set* of U ; thus $P(U) = \{S \mid S \subset U\}$. For example, if U has two elements, it has four different subsets which are the four elements of $P(U)$. Explicitly, $P(\{1, 2\}) = \{\{1, 2\}, \{1\}, \{2\}, \emptyset\}$. Here \emptyset is the empty set (a subset of every set, as above).

EXERCISES

- For subsets R , S , and T of a set U , establish the following identities:
 - $R \cap S = S \cap R$, $R \cap (S \cap T) = (R \cap S) \cap T$.
 - $R \cup S = S \cup R$, $R \cup (S \cup T) = (R \cup S) \cup T$.
 - $(R \cap S)' = R' \cup S'$, $(R \cup S)' = R' \cap S'$.
 - $S \cap (S \cup T) = S$, $S \cup (S \cap T) = S$.
- Show that any one of the three conditions $S \subset T$, $S \cap T = S$, and $S \cup T = T$ on the sets S and T implies both of the others.
- For $S \subset U$, show that $S \cap S' = \emptyset$ and $S \cup S' = U$.
- List the elements of the sets $P(P(\{1\}))$ and $P(P(P(\{1\})))$.
- Show that a set of n elements has 2^n different subsets.
- If $m < n$, show that a set of n elements has $(n!)/(n-m)!(m!)$ different subsets of m elements each, where $m! = 1 \cdot 2 \cdot \cdots \cdot m$.

2. Functions

A *function* f on a set S to a set T assigns to each element s of S an element $f(s) \in T$, as indicated by the notation

$$s \mapsto f(s), \quad s \in S.$$

The element $f(s)$ may also be written as fs or f_s , without parentheses; it is the *value* of f at the *argument* s . The set S is called the *domain* of f , while T is the *codomain*. The *arrow* notation

$$f: S \rightarrow T \quad \text{or} \quad S \xrightarrow{f} T$$

indicates that f is a function with domain S and codomain T . A function is often called a “map” or a “transformation”.

To describe a particular function, one must specify its domain and its codomain, and write down its effect upon a typical (“variable”) element of its domain. Thus the squaring function $f: \mathbb{R} \rightarrow \mathbb{R}$ for the set \mathbb{R} of real numbers may be described in any of the following ways: As the function f with $f(x) = x^2$ for any real number x , or as the function $(-)^2$, where $-$ stands for the argument, or as the function which sends each $x \in \mathbb{R}$ to x^2 , or as the

function given by the assignment $x \mapsto x^2$ for $x \in \mathbb{R}$. We systematically use the *barred arrow* to go from argument to value of a function and the *straight arrow* $S \rightarrow T$ to go from domain to codomain.

Note that a letter such as f or g stands for a function, while an expression such as $f(x)$ or $g(x)$ stands for a value of that function for an element x of its domain. For example, in trigonometry the expression " $\sin x$ " stands for a number, so we speak not of "the function $\sin x$ " but of the function $\sin: \mathbb{R} \rightarrow \mathbb{R}$. By using a barred arrow, we can describe particular functions without naming them as f or g ; for example, $x \mapsto x^2 + 3x + 2$ for x real describes a function $\mathbb{R} \rightarrow \mathbb{R}$.

Two functions f and g are *equal* (in symbols, $f = g$) when they have the same domain, the same codomain, and the same value $f(s) = g(s)$ for each element s of this common domain. For example, the assignment $x \mapsto x + 2$ (add two) defines on the integers \mathbb{Z} a function $f: \mathbb{Z} \rightarrow \mathbb{Z}$; on the set \mathbb{R} of real numbers it also defines a function $g: \mathbb{R} \rightarrow \mathbb{R}$; these are *different* functions because they have different domains.

The *image* of a function $f: S \rightarrow T$ is the set $f(S)$ of all values $f(s)$ for $s \in S$; it is always a subset of the codomain of f .

For any set S , the *identity function* $1_S: S \rightarrow S$ is that function $s \mapsto s$ which maps each element s of S onto itself. Different sets have different identity functions. If S is a subset of U , the *insertion* $i: S \rightarrow U$ is that function on S to U which assigns to each element of S the same element, now in U . Note that "insertion" is a function $S \rightarrow U$ and "inclusion" a relation $S \subset U$; every inclusion relation gives rise to an insertion function.

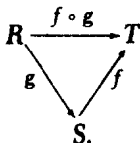
The *composite* $f \circ g = fg$ of two functions is the function obtained by applying them in succession; first g , then f —provided this makes sense; that is, provided the domain of f is the codomain of g . More formally, given the functions

$$g: R \rightarrow S, \quad f: S \rightarrow T,$$

their *composite* is the function $f \circ g: R \rightarrow T$ with values given by

$$(f \circ g)(r) = f(g(r)), \quad \text{all } r \in R. \quad (4)$$

This definition may be visualized by the "mapping diagram" displayed below:



To go from the set R directly to the set T by the composite $f \circ g$ is the same as going through S in two steps, the first by g and the second by f . We also express this fact by saying: This triangular diagram "commutes".

Composition of functions obeys the

Associative law: $(f \circ g) \circ h = f \circ (g \circ h)$,

whenever the composites involved are defined. This is obvious intuitively; both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ have the effect of applying first h , then g , and finally f , in that order. Formally, given $h: P \rightarrow R$, $g: R \rightarrow S$, and $f: S \rightarrow T$, both triple composites $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are functions on P to T , while the first composite assigns to each $p \in P$ the value

$$[(fg)h]p \stackrel{(fg)h}{=} (fg)(hp) \stackrel{fg}{=} f(g(hp)) \stackrel{g}{=} f((gh)p) \stackrel{f(gh)}{=} [f(gh)]p;$$

here each step depends on applying the definition (4) of composition to the composite indicated below the equality symbol for that step. By the definition of equality for functions, this proves the associative law $(fg)h = f(gh)$: $P \rightarrow T$. Note that here (and often later) it is convenient to omit the symbol " \circ " in $f \circ g$ and the parentheses in $h(p)$ or $(gh)p$.

Under composition each function $f: S \rightarrow T$ obeys the

Identity law: $f \circ 1_S = f = 1_T \circ f: S \rightarrow T$.

To prove the first equality, note by (4) that $(f1_S)s = f(1_S s) = fs$ for all $s \in S$; hence, $f1_S = f$, by the definition of equality for functions. The second equality is proved similarly.

A function $r: S \rightarrow T$ is said to be a *restriction* of a function $f: U \rightarrow V$ when $S \subset U$, $T \subset V$, and $r(s) = f(s)$ for each $s \in S$. (One also then says that f is an *extension* of the function r .) For example, given a subset $S \subset U$, the insertion $i: S \rightarrow U$ is a restriction of the identity $1_U: U \rightarrow U$.

Certain useful special types of functions will now be defined.

A function $f: S \rightarrow T$ is *injective* or an *injection* when $s_1 \neq s_2$ in S implies $fs_1 \neq fs_2$ in T ; that is, when f carries distinct elements of its domain to distinct elements of its codomain. For example, every insertion is an injection. A function $h: S \rightarrow T$ is *surjective* or a *surjection* when its image is the whole codomain T ; that is, when to each $t \in T$ there exists at least one $s \in S$ with $hs = t$. Finally, a *bijection* $b: S \rightarrow T$ is a function which is both an injection and a surjection; thus b is *bijective* if and only if to each $t \in T$ there is exactly one element $s \in S$ with $bs = t$. The notation \cong , as in $b: S \cong T$, indicates that b is a bijection of S to T .

For example, among the functions $\mathbf{Z} \rightarrow \mathbf{Z}$, the function $n \mapsto (-n)$ is a bijection, the function $n \mapsto 2n$ is an injection but not a surjection, and the function $n \mapsto n^2$ is neither an injection nor a surjection.

Again, for example, if \mathbf{R}^+ is the set of all nonnegative real numbers, the squaring function $g: \mathbf{R} \rightarrow \mathbf{R}^+$ given by $g(x) = x^2$ is surjective, because every nonnegative real is the square of some real number. However, the squaring function $f: \mathbf{R} \rightarrow \mathbf{R}$ with $f(x) = x^2$ and codomain *all* the real numbers is *not*

surjective. These two squaring functions (though they have the same values) count as different functions because they have different codomains: Whether or not a function is a surjection *depends on its codomain*.

There is another parallel terminology for these ideas:

Injection $S \rightarrow T =$ "one-one" map of S "into" T ;

Surjection $S \rightarrow T =$ map of S *onto* T ;

Bijection $S \rightarrow T =$ one-one map of S onto T ,

or, in the last case, a "one-one correspondence" of S to T . The older terminology (that to the right) will not be used in this book.

Any function f can be written as a composite $f = g \circ h$, where g is injective and h surjective. Indeed, if $f: S \rightarrow T$ has image $U \subset T$, its restriction $r: S \rightarrow U$ is a surjection, the insertion $i: U \rightarrow T$ is an injection, and f itself is the composite $f = i \circ r$.

Certain functions have "inverses". Suppose that $g: T \rightarrow S$ and $f: S \rightarrow T$, so that the composite $f \circ g$ is defined. If this composite is the identity $1_T = f \circ g$, call f a *left inverse* of g and g a *right inverse* of f . When the composites in both orders are identities, so that $f \circ g = 1_T$ and $g \circ f = 1_S$, call f a *two-sided inverse* of g (and hence g a two-sided inverse of f).

THEOREM 1. *A function with non-empty domain is an injection if and only if it has a left inverse. A function is a surjection if and only if it has a right inverse.*

Proof: Suppose first that $g: T \rightarrow S$ has a left inverse $f: S \rightarrow T$. Then $fg = 1_T$, so $g(t_1) = g(t_2)$ implies $t_1 = fg(t_1) = fg(t_2) = t_2$. Therefore, g is injective. Conversely, suppose that $g: T \rightarrow S$ is injective with domain $T \neq \emptyset$; pick some $t_0 \in T$. Since g is injective, there is to each $s \in S$ at most one t with $s = g(t)$; hence, a function $f: S \rightarrow T$ is defined by

$$\begin{aligned} f(s) &= \text{that } t \text{ with } g(t) = s, & \text{when } s \in \text{image } (g), \\ &= t_0, & \text{otherwise.} \end{aligned}$$

This function f sends each $g(t)$ "back where it came from"; so $f(g(t)) = t$ for every t . This states that $f \circ g = 1_T$, so f is the desired left inverse for g .

Note, however, that an injection g which is not a bijection will have in general many left inverses; for example, one for each choice of the t_0 above.

It remains to prove the second half of the theorem, concerning surjections. Suppose first that a function $f: S \rightarrow T$ has a right inverse g . Now $1_T = f \circ g$ means that $t = f(gt)$ for all t , so each $t \in T$ is in the image of f , and f is surjective, as required. Conversely, suppose that $f: S \rightarrow T$ is surjective; this means that to each $t \in T$ there is at least one $s \in S$ with $f(s) = t$. Choose

one such s for each t and define $g: T \rightarrow S$ by letting $g(t)$ be the chosen s . Then $f(g(t)) = t$, so $f \circ g = 1_T$, and g is the desired right inverse. This completes the proof of the theorem.

Note: This proof depends on making a (possibly) infinite number of choices (one $s \in S$ with $f(s) = t$ for each $t \in T$). In an axiomatic treatment of set theory, when all the operations on sets are derived from a complete list of formal axioms on the membership relation $x \in S$, one of the axioms states that such a set of choices can be made. This axiom, called the "axiom of choice", states that to each set \mathcal{T} whose elements are disjoint nonvoid sets, there exists a set C such that each $C \cap S$, for $S \in \mathcal{T}$, has exactly one element. This axiom is equivalent to the assumption that every surjection has a right inverse (Exercise 11).

COROLLARY. *The following properties of a function $g: T \rightarrow S$ are equivalent:*

- (i) g is a bijection.
- (ii) g has both a left inverse f and a right inverse h .
- (iii) g has a two-sided inverse.

When this is the case, any two inverses (left, right, or two-sided) of g are equal. This unique inverse of g (written g^{-1}) is bijective, and satisfies

$$(g^{-1})^{-1} = g. \quad (5)$$

Proof: First suppose that $T \neq \emptyset$ and $S \neq \emptyset$, so the theorem can be used. Since a bijection is both surjective and injective, the theorem at once gives the equivalence of (i) and (ii). As for (iii), any two-sided inverse is trivially both a left and a right inverse; thus (iii) implies (ii). Conversely, (ii) implies

$$f = f \circ 1_S = f \circ (g \circ h) = (f \circ g) \circ h = 1_T \circ h = h,$$

which means that $f = h$ is a two-sided inverse for g ; hence (ii) gives (iii). This argument also shows that any left inverse f of g must equal any right inverse h ; this is the next clause of the corollary. Finally, the inverse $f = h$ of g has g for a two-sided inverse, hence, it is also bijective and has g as its inverse. This is the conclusion (5) of the corollary.

Only the (uninteresting) case of the corollary when T or S is empty remains: Now a function $g: \emptyset \rightarrow S$ with empty domain must assign to each element of \emptyset an element of S . But there are no elements of the empty set \emptyset , so there is exactly one function $\emptyset \rightarrow S$ (namely, the one which involves no assignments). If $S \neq \emptyset$, this function g is not a bijection; on the other hand, there can be no function $S \rightarrow \emptyset$, so g has no inverses of any sort. Thus the corollary holds in this case. If $S = \emptyset$ it holds trivially, and hence in all cases.

Note incidentally that the function $g: \emptyset \rightarrow S$ with $S \neq \emptyset$ is injective but has no left inverse.

If $g: T \rightarrow S$ and $k: S \rightarrow R$ are both bijections, so is their composite $k \circ g$; its inverse is given by

$$(k \circ g)^{-1} = g^{-1} \circ k^{-1} \quad (\text{reverse the order}). \quad (6)$$

Indeed, $k^{-1} \circ k = 1_S$ and $g^{-1} \circ g = 1_T$, so, by the associative law,

$$g^{-1}k^{-1}kg = g^{-1}1_Sg = g^{-1}g = 1_T,$$

and $g^{-1}k^{-1}$ is a left inverse for kg . A similar calculation shows that it is also a right inverse, hence the conclusion and (6).

Conventions on functions differ. In this discussion (as elsewhere in this book), we have written each function to the left of its argument, as in $f(s)$ —and as is customary in analysis and topology. In consequence, the composite $f \circ g$ has meant *first* apply g , *then* apply f . Functions may also be written to the right of their arguments; then a composite has the opposite meaning.

E X E R C I S E S

1. If $S = \{0, 1\}$ is a set with exactly two elements, exhibit all functions $S \rightarrow S$ and classify them as injective, surjective, bijective, or none of these.
2. If $f \circ g$ is defined and both f and g have left inverses, show that $f \circ g$ has a left inverse.
3. Show that the composite of two surjections is a surjection, and similarly for injections.
4. If f is a bijection and $f \circ g$ is defined, show that g is an injection if and only if $f \circ g$ is, and a surjection if and only if $f \circ g$ is one.
5. With N the set of nonnegative integers, show that the function $f: N \rightarrow N$ given by $n \mapsto n^2$ has no right inverse, and exhibit explicitly two left inverses.
6. If f is injective, while both $f \circ g$ and $f \circ g'$ are defined, show that $f \circ g = f \circ g' \Rightarrow g = g'$.
7. Find an analog of Exercise 6 for surjections.
8. For any function $f: S \rightarrow T$ with $S \neq \emptyset$, construct a function $h: T \rightarrow S$ with $fhf = f$. Deduce from this the results of Theorem 1.
9. Show that a function which has a unique right-inverse is necessarily bijective.
10. Prove the Corollary of Theorem 1 without using the axiom of choice.
- *11. Assuming that every surjection has a right inverse, prove the axiom of choice as stated in the text. (*Hint:* Let U be the union of all $S \in \mathcal{F}$, define $f: U \rightarrow \mathcal{F}$ by $f(u) = S$ when $u \in S$, and show f surjective.)

3. Relations and Binary Operations

To treat functions of two variables or relations between two variables we use "ordered pairs". The *ordered pair* consisting of two elements, s and t , in that order, is written (s, t) . The equality of two ordered pairs is defined by the rule

$$(s, t) = (s', t') \Leftrightarrow s = s' \text{ and } t = t'.$$

The *cartesian product* $S \times T$ of two sets S and T is defined to be the set of all ordered pairs (s, t) of elements from S and T , respectively. Thus

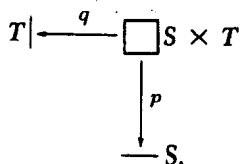
$$S \times T = \{(s, t) | s \in S, t \in T\}.$$

Thus, if \mathbf{R} is the set of all real numbers, $\mathbf{R} \times \mathbf{R}$ is the set of all ordered pairs (x, y) of real numbers; in other words, $\mathbf{R} \times \mathbf{R}$ is just the set of all cartesian coordinates of points in the plane (relative to given coordinate axes).

Any cartesian product $S \times T$ may be "projected" onto its "axes", S and T :

$$S \xleftarrow{p} S \times T \xrightarrow{q} T.$$

These *projections* are the functions p and q defined by $p(s, t) = s$ and $q(s, t) = t$, as in the diagram



We call this the *cartesian-product diagram*.

Note the bijection $S \times T \cong T \times S$ given by $(s, t) \mapsto (t, s)$.

Ordered triples may be described in terms of ordered pairs. Given r, s , and t , define the *ordered triple* (r, s, t) to be $(r, (s, t))$. Write $R \times S \times T$ for the set $R \times (S \times T)$ of all triples $(r, (s, t))$ for $r \in R, s \in S$, and $t \in T$, and note that the assignment $(r, (s, t)) \mapsto ((r, s), t)$ is a bijection $R \times (S \times T) \cong (R \times S) \times T$. Ordered "quadruples" (r, s, t, u) and the like are defined similarly.

One may also form the cartesian product of functions. Given two functions $u: S \rightarrow S'$ and $v: T \rightarrow T'$, their *cartesian product* is the function $u \times v: S \times T \rightarrow S' \times T'$ defined by $(u \times v)(s, t) = (us, vt)$.

The cartesian product is useful in describing the usual functions of two or more arguments; a function F with two arguments $s \in S$ and $t \in T$ and with values in a set W is a function

$$F: S \times T \rightarrow W$$

on the cartesian product $S \times T$ to the set W . Such a function F assigns to each ordered pair $(s, t) \in S \times T$ a value $F(s, t) \in W$.