FUNDAMENTAL STRUCTURES OF ALGEBRA AND DISCRETE MATHEMATICS

STEPHAN FOLDES

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Brisbane / Toronto / Singapore

This page intentionally left blank

FUNDAMENTAL STRUCTURES OF ALGEBRA AND DISCRETE MATHEMATICS

This page intentionally left blank

FUNDAMENTAL STRUCTURES OF ALGEBRA AND DISCRETE MATHEMATICS

STEPHAN FOLDES

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Brisbane / Toronto / Singapore

This text is printed on acid-free paper.

Library of Congress Cataloging in Publication Data:

Foldes, Stephan, 1951–
Fundamental structures of algebra and discrete mathematics/by Stephan Foldes.
p. cm.
"A Wiley-Interscience publication."
Includes bibliographical references and index.
ISBN 0-471-57180-6 (cloth)
1. Algebra. I. Title.
QA155.F65 1993
512'.02–dc20 93-8787

10 9 8 7 6 5 4 3 2 1

In token of my respect and gratitude
this book is dedicated to
Ilonka and Sándor Szász

This page intentionally left blank

CONTENTS

This page intentionally left blank

ILLUSTRATIONS

PREFACE

This book is about the algebraic notion of "structure". In mathematical thinking, a structure crystallizes whenever attention is focused either on combining elementary objects of some kind to form other objects of a similar kind, such as adding numbers to form new numbers, or on relating objects to each other, such as comparing numbers by magnitude. Instead of numbers, two points in space may be combined to define a line, a point and a line may be combined to define a plane, and these geometric objects may also be linked by relations such as inclusion and parallelism. Numbers represent a significant abstraction from whatever is being weighed or enumerated, and straight lines miss much of the reality of land surveying. However, algebra involves a second shift in interest, from the things combined to the ways of combination.

Numbers with addition constitute the historical archetype of algebraic structure. If negatives are included, we have a group; if not, we have a semigroup. If multiplication is taken into consideration as well as addition, then a more complex structure called a ring arises. There is the ring of integer numbers, and the ring of rational numbers, and so on. Most important for the algebraist is the realization that there are rings consisting of objects that are not numbers at all; objects that can be added and multiplied and that obey rules such as $a(b + c) = ab + ac$.

This volume presents a basic theory of groups and rings and other algebraic structures. Like most algebra texts, it has a chapter on

fields and one on vector spaces. Like some more recent texts, it includes lattices and universal algebras. The classical number systems Z, Q, R and C find their *raison d'être* in abstract algebra, and not the other way round.

The machinery is based on sets, order relations, and closure operators. All mathematical objects are defined in terms of sets. The entire theory is derived from nine set-theoretical axioms. Sets can also be viewed as the simplest kind of all structures. They are the subject of Chapter I.

There are two reasons to study order relations in algebra. First, more or less obvious order relations present in various algebraic structures provide simple explanations of what is going on. Second, the study of ordered sets, as a kind of structure, can be undertaken in the same spirit as the study of structures with a law of composition. To a lesser extent the same is true for graphs.

Binary operations more general than those of groups are needed to discuss ring multiplication, lattices, and word concatenation. Partial binary operations are needed for categories. However, in the last chapter on categories we do not enter into any generalization that is not directly relevant to the material in the preceding chapters, despite (or rather because of) all the new algebra that could be thus presented.

The fundamental role that closure operators play in algebra led us to view matroids and topological spaces as structures of an algebraic nature. This is why we devoted a separate chapter to each, rather than confining them to subsidiary treatment under "geometric lattices" and "filters in Boolean lattices."

Students of algebra and researchers in other areas will find in this book an introduction to, or a clarification of, the basic theories of the twelve kinds of structures. A comprehensive exposition of each particular theory is not the aim of this text. Rather, we seek to identify essentials and to describe interrelationships between particular theories. We hope that the specialist of commutative algebra will find matroids worth the reading and that the student of discrete mathematics will find special relativity close to his or her own field.

The material is self-contained. The reader need not know any mathematical definitions, results, or methods. However, the pace and density of exposition corresponds to those of graduate texts.

Selected advanced results are derived from weak rather than strong hypotheses, whenever this is compatible with the objective of simplicity. Also, several classical concepts are introduced relatively late, in order to demonstrate the simplicity of certain results established without the use of these concepts. Thus, Zermelo's Theorem is proved before set intersections are introduced, elementary group theory (including Lagrange's Theorem) is developed before the theory of integers, and a simple Galois theory is presented without calling on vector space dimensions. And if some major algebraic concept fails to appear altogether, the reader may conclude that it is not required for any of the theorems included in this short volume.

Throughout the text the student is frequently prompted to perform exercises of verification and to explore examples. These integrated exercises are indispensable for any reader not yet familiar with the theory. At the end of each section, there are additional numbered exercises from which to select. Many of these are open-ended questions in the sense that while a satisfactory answer can be given without much difficulty, there is ample room for better and more complete answers. (An exact science mathematics may well be, but mathematical research is not more deterministic than any other intellectual endeavor.)

Each chapter builds on key notions introduced in previous chapters. However, if you are already conversant with some of the structures, then you may go directly to selected chapters or even sections and use the index whenever you suspect a divergence between your definitions and ours.

In a course or seminar, the book should be used as reading material before or after verbal presentations. If the syllabus is limited to certain chapters only, the instructor should summarize for the students the definitions needed from the excluded chapters. In a seminar, we recommend the use of a Socratic approach, with analysis of examples and blackboard exercises, to probe the students' understanding of how constructs relate to theorems and structures to each other. The ultimate object of probing, however, is not progress in learning a science, but the tools and concepts of science itself. It is in the spirit of such questioning that the book was written.

This page intentionally left blank

CHAPTER I

SETS

1. ELEMENTARY CONSTRUCTIONS AND AXIOMS

The ability to think about collections of objects with precision and without ambiguity is indispensable in mathematics. Indeed, this is so in any exact science. Students of the physical world care about and count the collection of atoms in a given portion of matter, and chemists concern themselves with the collection of atoms forming a molecule, distinguishing various compounds according to what kind of molecules they contain and in what proportions. Biology in turn makes use of physics and chemistry and describes aggregates of millions of cells forming a tissue. Living and extinct creatures are classified into collections called species and their subcollections called varieties, and resembling species are grouped into families and kinds. Economists conceptualize and measure physical quantities of edible or otherwise useful goods, taking stock of grain, cattle, and money supply, distinguishing raw materials from work-in-process inventories, and discussing such issues as whether home-baked cakes should be included in or excluded from the gross national product. Linguists divide the collection of all words into subcollections such as nouns, pronouns, verbs, and adverbs, and they study small groups of words called sentences as they relate to the former subcollections.

No other science, however, relies as much on the conceptual manipulation of collections as mathematics.

It was already realized by the reflective Greeks of pre-Christian times that some restraint must be exercised in talking about collections. The liar's paradox of Crete goes as follows. Let collection T be the collection of all true sentences uttered on the island of Crete, and let F be the collection of untrue (false) sentences. Then let a mathematician take a boat to Crete and upon disembarking declare: "This very sentence I am pronouncing at this moment belongs to F." Despite the apparent rigor employed in defining T and F, the rules of logic seem to break down. If the mathematician's declaration is true, then the sentence being pronounced does indeed belong to F, and therefore it is false by definition of F. But if the declaration is false, then it must belong to F by definition, and it therefore becomes a true statement. Modern mathematics deals with the paradox by imposing very stringent rules on how collections should be defined. Requiring a precise definition of any mathematical object before making statements about that object is what lends mathematics its reliability. (Look at the contractual text of an insurance policy for an analogy.) Although very restrictive, the rules of definition still allow all usual mathematical objects to be defined in terms of collections. Indeed, the space of three-dimensional geometry will be a collection of vectors, of which points, lines, and planes will be subcollections. Moreover, the vectors themselves will be defined as collections, and the numbers 0, 1, 2, etc., will be formally defined as collections of a most particular kind.

The entire body of mathematical science can be viewed as a theory about collections called *sets*. By using the technical word "set," mathematicians simply indicate that they are talking about a collection and that they strongly believe that they know what they are talking about. Accordingly, the mathematician may wish to avoid the unregulated word "collection." To what extent this suffices to exempt mathematics from the fundamental uncertainty that affects human knowledge is open to debate. First, every mathematical discourse has a small number of primitive concepts that are not defined rigorously but are used in the formal definition of more elaborate concepts. Second, a few simple mathematical propositions are based on belief and observation rather than proof, yet they serve as the very founda-

tion of all further theory. Having disclosed these risk factors, let us proceed.

An object x belonging to a set S is termed an *element*, or *member*, of S, in symbols $x \in S$. If the object x does not belong to S, we write $x \notin S$. Sets are completely determined by their elements, i.e., if two sets A and B have the same elements, then the two sets are the same. In this case we write $A = B$. If A and B are not the same, then we write $A \neq B$. If all elements of A are also elements of B, then A is called a *subset* of B, in symbols $A \subseteq B$. We also say that B is a *superset* of A and write $B \supseteq A$. The negation of $A \subseteq B$ is written $A \not\subseteq B$. Trivially, every set B has at least one subset, for $B \subseteq B$. A subset A of B is a *proper subset* if $A \neq B$, and we then write $A \subset B$. The following axiomatic propositions are adopted, without proof, entirely on the basis of their intuitive plausibility.

(A1) Empty Set Axiom. *There is a set $\emptyset$ which has no element.*

The set $\emptyset$ is called the *empty set* (or *null* or *void* set). Clearly $\emptyset \subseteq A$ for every set A.

(A2) Subset Axiom. *If A is any set, then those elements x of A that satisfy some given condition or possess a given property form a set.*

To designate "the set of those elements x of A that satisfy a certain condition," we usually write $\{x \in A : x \text{ satisfies a certain condition}\}$. This is of course a subset of A. In practice, the original superset specification "$x \in A$" is often indicated in other, less explicit ways.

(A3) Power Set Axiom. *For any set A, there is a set $\mathcal{P}(A)$, whose elements are all the subsets of A.*

The set $\mathcal{P}(A)$ is called the *power set* of A. Occasionally we write simply $\mathcal{P}A$ instead of $\mathcal{P}(A)$.

Since $A \subseteq A$, we have $A \in \mathcal{P}(A)$, and it follows from (A2) that $\mathcal{P}(A)$ has a subset containing A as unique element. This subset of $\mathcal{P}(A)$ will be denoted by $\{A\}$ and called a *singleton* set. We have $\{A\} = \{x \in \mathcal{P}(A) : x = A\}$.

(A4) Pair Axiom. *If A and B are sets, then there is a set $\{A, B\}$ that has both A and B as elements but has no other elements.*

The set $\{A,B\}$ is called a *pair*. Note that $\{A,B\} = \{B,A\}$. If $A = B$, then the pair is a singleton, $\{A,B\} = \{A\} = \{B\}$.

(A5) Union Axiom. *Given any set A whose members are sets, there is a set $\cup A$ whose elements are the elements of the members of A.*

The set $\cup A$ is called the *union* of A (or of the members of A), and it is also denoted by $\bigcup_{X \in A} X$. The union of a pair $\{A,B\}$ is usually denoted by $A \cup B$ and called the *union* of A and B. Thus if we let $I = \{A,B\}$, then

$$A \cup B = \cup I = \bigcup_{X \in I} X$$

These few axioms immediately allow us to ask and answer a meaningful question that is quintessentially algebraic. Suppose we have sets A, B, and C. Taking the union of $A \cup B$ and C we get some set $(A \cup B) \cup C$. Proceeding differently, taking the union of A with $B \cup C$, we get some set $A \cup (B \cup C)$. But are not $(A \cup B) \cup C$ and $A \cup (B \cup C)$ the same? Yes indeed, because it is easy to verify that they have the same elements. It is time to state the first theorem of algebra:

Proposition 1 *Let A, B, and C be sets.*

(i) $(A \cup B) \cup C = A \cup (B \cup C)$ (*associative law*)
(ii) $A \cup B = B \cup A$ (*commutative law*)
(iii) $A \cup A = A$ (*idempotent law*)

Proof. Associativity has just been observed. Commutativity follows from the earlier made observation that $\{A,B\} = \{B,A\}$, which is often referred to by saying that the pair $\{A,B\}$ is "not ordered." Finally, idempotence is obvious by definition of the union. □

Using axioms (A1) to (A5), we can define a great variety of sets. The reader should verify which particular axioms need to be called upon to construct the following specific examples:

The empty set $\emptyset$.
The singleton $\{\emptyset\}$.
The pair $\{\emptyset, \{\emptyset\}\}$.

The power set $\mathcal{P}(\{\emptyset,\{\emptyset\}\})$.
The power set of the above, $\mathcal{PP}(\{\emptyset,\{\emptyset\}\})$.
The power set of the above.

And so forth without end. What is remarkable here is that each of these examples is a *set of sets*, a set whose elements are themselves sets. Indeed we shall only consider sets of sets in this volume. It is the author's view, adopted in this book at least, that in mathematics we need not and should not speak about sets of atoms, molecules, animals, or true or false sentences unless these various objects can be precisely defined as sets themselves. In some cases this may be done meaningfully, such as in theoretical physics, or in mathematical logic where sentences can be defined as proper mathematical objects themselves, i.e., as sets. Neither should we speak about the set of even or odd numbers until these numbers have been defined as sets: this will be done in a while.

It was pointed out that the elements of a pair $\{a,b\}$ are "not ordered," $\{a,b\} = \{b,a\}$. The *ordered pair* $\langle a,b\rangle$ is defined by

$$\langle a,b\rangle = \{\{a\},\{a,b\}\}$$

and it has the desired property that

$$\langle a,b\rangle = \langle c,d\rangle \qquad \text{if and only if both} \quad a = c \quad \text{and} \quad b = d$$

Thus $\langle a,b\rangle = \langle b,a\rangle$ if and only if $a = b$. The reader should verify that $\langle a,b\rangle$ is a subset of $\mathcal{PP}(a \cup b)$. If a and b are elements of sets A and B, respectively, then

$$a \cup b \subseteq (\cup A) \cup (\cup B)$$

and thus $\langle a,b\rangle$ is also a subset of

$$\mathcal{PP}((\cup A) \cup (\cup B))$$

Therefore those ordered pairs $\langle x,y\rangle$ for which $x \in A$ and $y \in B$ form a subset $A \times B$ of

$$\mathcal{PPP}((\cup A) \cup (\cup B))$$

called a *Cartesian product*. A *function*, *map*, or *mapping from A to B* is then any subset f of $A \times B$ such that for every $x \in A$ there is a unique $y \in B$ with

$$\langle x,y\rangle \in f$$

The set A is called *the domain*, B *a codomain* of f. (A function can have many different codomains, for if $B \subseteq B'$, then $A \times B \subseteq A \times B'$, and thus every function from A to B is also a function from A to B'.) We use the shorthand $f : A \to B$ for "a function f from A to B." For $x \in A$, the unique element y of B such that $\langle x, y \rangle \in f$ is called the *image of* x *by* f, or the *value of* f *on* x, and it is denoted by $f(x)$. It is also said that f *associates* $f(x)$ with x, f *maps* x to $f(x)$, or $f(x)$ is obtained by *applying* f to x. The set of all functions from A to B is denoted by B^A; it is a subset of $\mathcal{P}(A \times B)$. For reasons to be seen later, we say that B^A is obtained from B and A by *set exponentiation*. For $f \in B^A$ and $S \subseteq A$ the function

$$\{\langle x, f(x) \rangle : x \in S\}$$

is called the *restriction* of f to S. It is a function from S to B, and it is denoted by $f|S$. We also say that f is an *extension* of $g = f|S$ to A, or that f *extends* g.

Observe that two functions $f, g \in B^A$ are identical if

$$f(x) = g(x) \qquad \text{for every} \quad x \in A$$

A function is thus completely determined by its values on the various elements of its domain, and usually that is how functions are specified.

Informally, a function $A \to B$ is often thought of as a "rule," "procedure," or "machine" that, given any input $x \in A$, "allows us to find" or "produces" an element $f(x) \in B$. Many functions seen in mathematics appear in fact to fit this notion. However, many other functions, perhaps not "seen" but existing nevertheless, have nothing to do with computational procedures. (This issue is of great importance in mathematical philosophy and logic and of practical relevance in computer science. The interested reader is referred to the theory of recursive functions and to the theory of computational complexity.)

For every set A the function $i : A \to A$ defined by

$$i = \{\langle x, x \rangle \in A \times A : x \in A\}$$

[or equivalently by $i(x) = x$ for $x \in A$] is called the *identity function* on A, often denoted by id_A. In a nonempty set B let $b \in B$. Assume also that $A \neq \emptyset$. The function $c = A \times \{b\}$ from A to B is said

to be *constant* because

$$c(x) = b \qquad \text{for all} \quad x \in A$$

An arbitrary map $f : A \to B$ is said to be *constant on a subset* S of its domain A if the restriction $f|S$ is constant.

The *image of a set* $S \subseteq A$ by a function $f : A \to B$ is the set

$$\{y \in B : y = f(x) \text{ for some } x \in S\}$$

It is denoted by $f[S]$. The *inverse image of a set* $T \subseteq B$ is

$$\{x \in A : f(x) \in T\}$$

It is denoted by $f^{\text{inv}}[T]$. The *image*, or *range*, of the function $f : A \to B$ is $f[A]$; it is denoted by $\text{Im} f$. If $\text{Im} f = B$, then f is said to be *surjective to* B (or a function *onto* B). All identity functions are surjective onto their own domains. On the other hand, the image of a constant function is a singleton, and therefore a constant function $c : A \to B$ is not surjective onto B unless the codomain B is a singleton.

A function $f : A \to B$ is *injective* (or an *injection*) if there are no distinct elements $x \neq x'$ of A with $f(x) = f(x')$. All identity functions are injective. On the other hand, a constant function is not injective unless its domain is a singleton. (The reader should verify this.)

An injective function surjective onto a codomain B is called *bijective* (or a *bijection*) *to* B. All identity functions are bijective to their own domains. For a nontrivial example, let S be any set and let the *complementation function* $f : \mathcal{P}(S) \to \mathcal{P}(S)$ be defined by

$$f(A) = \{x \in S : x \notin A\} \qquad \text{for every} \quad A \in \mathcal{P}(S)$$

This complementation function is bijective from $\mathcal{P}(S)$ to $\mathcal{P}(S)$.

There are two facts that we should take note of at this juncture. First, for any set S there is an injection $f : S \to \mathcal{P}(S)$. Indeed, f can be defined by $f(x) = \{x\}$ for all $x \in S$, i.e.,

$$f = \{\langle x, \{x\}\rangle : x \in S\}$$

Second, let us prove that there is no injection $g : \mathcal{P}(S) \to S$ from the power set of S into S. Suppose that such a g exists: we shall derive a contradiction. Consider those subsets A of S for which $g(A) \notin A$. Let F be the set of the corresponding elements $g(A)$,

$$F = \{x \in S : x = g(A) \text{ for some } A \subseteq S \text{ such that } g(A) \notin A\}$$

If we had $g(F) \notin F$, then by letting $A = F$, it follows from the definition of F that $g(A)$ belongs to F, i.e., $g(F) \in F$. And if $g(F) \in F$, then $g(F) = g(A)$ for some $A \subseteq S$ such that $g(A) \notin A$, again referring to the definition of F verbatim. Since g is supposed to be injective, $g(F) = g(A)$ implies $F = A$. But then $g(F) \in F$ and $g(A) \notin A$ are contradictory, proving the absurdity of the alleged existence of an injective g. This argument is inspired by the liar's paradox. However, what is reduced to absurdity here is not the universal dichotomy of truth and falsehood but merely the possibility of injecting $\mathcal{P}(S)$ into S. The argument is indeed a domesticated variety of the liar's paradox of Crete.

If $f : A \to B$ and $g : B \to C$ are functions, then the *composition* $g \circ f$ is the function from A to C defined by

$$(g \circ f)(x) = g(f(x)) \qquad \text{for all} \quad x \in A$$

i.e., $g \circ f$ is the set of all ordered pairs $\langle x, z \rangle$ such that for some $y \in B$,

$$f(x) = y \qquad \text{and} \qquad g(y) = z$$

Occasionally we shall write simply gf instead of $g \circ f$. The reader can see that the composition of two injective functions is injective and the composition of surjective functions is surjective. Hence, the composition of bijections is bijective. Observe further that a function $f : A \to B$ is bijective to B if and only if the set

$$\{\langle y, x \rangle \in B \times A : \langle x, y \rangle \in f\}$$

is itself a function from B to A. Denoting this new function by f^*, we have $f^* \circ f = id_A$ and $f \circ f^* = id_B$, and f^* is called the *inverse* of f. The reader should verify that f^* itself is a bijection from B to A, having in turn f as its inverse. Moreover, a function $f : A \to B$ is bijective if and only if there is a function $g : B \to A$ such that

$$g \circ f = id_A \qquad \text{and} \qquad f \circ g = id_B$$

and in that case g must coincide with f^*. For all $T \subseteq B$ we have

$$f^*[T] = f^{\text{inv}}[T]$$

Note, however, that $f^*[T]$ is only defined for bijective f, while $f^{\text{inv}}[T]$ is always defined.

The following proposition, of constant use in mathematics, anticipates the subject of the last chapter on category theory:

Proposition 2 *Let $f : A \to B$, $g : B \to C$, and $h : C \to D$ be functions.*

(i) $h \circ (g \circ f) = (h \circ g) \circ f$ *(associative law)*
(ii) $id_B \circ f = f$ *and* $f \circ id_A = f$ *(neutrality of the identities)*

Proof. Let $x \in A$. The image of x by $h \circ (g \circ f)$ is given by h applied to

$$(g \circ f)(x) = g(f(x))$$

i.e., it is $h[g(f(x))]$. But the image of x by $(h \circ g) \circ f$ is $h \circ g$ applied to $f(x)$, i.e., $h[g(f(x))]$ again. Thus $(h \circ g) \circ f$ and $h \circ (g \circ f)$ take the same value on every $x \in A$, and therefore they are identical functions. The neutrality of the identities can be verified by the reader. □

Here is now an early result in equational algebra:

Proposition 3 *Let f be a bijection of a set A into itself. Then for any function $g : A \to A$ there is a unique function $x : A \to A$ satisfying the equation*

$$g = f \circ x \tag{1}$$

There is also a unique function y satisfying

$$g = y \circ f \tag{2}$$

Proof. Since f^* is the inverse of f, $x = f^* \circ g$ is a solution of (1), because

$$f \circ (f^* \circ g) = (f \circ f^*) \circ g = id_A \circ g = g$$

Also, if x is any function satisfying $g = f \circ x$, then

$$f^* \circ g = f^* \circ (f \circ x) = (f^* \circ f) \circ x = id_A \circ x = x$$

This proves that x must be equal to $f^* \circ g$ and cannot be any other function. The unique solvability of (2) is shown similarly. □

A set A is said to be *equipotent* to B if there is a bijection from A to B. We shall then write $A \simeq B$.

Proposition 4 *Let A, B, and C be sets.*

(i) $A \simeq A$ (*reflexivity*)
(ii) *if* $A \simeq B$, *then* $B \simeq A$ (*symmetry*)
(iii) *if* $A \simeq B$ *and* $B \simeq C$, *then* $A \simeq C$ (*transitivity*)

Proof. Reflexivity results from the fact that the identity id_A is a bijection. Symmetry follows from the observation, made earlier, that the inverse of any bijection is a bijection. Finally, since the composition of two bijections is again a bijection, we have transitivity. □

A bijection f from a set A to a set B establishes what is often called a one-to-one correspondence between the elements of A and those of B. The elements of A and B are matched into ordered pairs $\langle a,b\rangle \in f$, with each element a of A being matched to a unique $b \in B$, and each $b \in B$ corresponding to a unique element of A. It is then tempting to say that A and B have the same number of elements. While we must refrain from using the word "number" until it is defined, this is actually what the term "equipotent" is meant to convey. The following result on set exponentiation may then be thought of as the first theorem of arithmetic:

Proposition 5 *For any sets A, B, and C we have*

$$(C^B)^A \simeq C^{B\times A}$$

Proof. Define a function F from $(C^B)^A$ to $C^{B\times A}$ as follows. If $f \in (C^B)^A$, then for every element $a \in A$, $f(a)$ is a function from B to C. Let then $\overline{f}$ be a function from $B \times A$ to C defined by

$$\overline{f}(\langle b,a\rangle) = f(a)(b)$$

Let $F(f)$ be defined as $\overline{f}$. The function F is injective for if $F(f) = F(g)$, then

$$f(a)(b) = g(a)(b) \qquad \text{for all} \quad \langle b,a\rangle \in B \times A$$

i.e., for $a \in A$ fixed the functions $f(a)$ and $g(a)$ from B to C are the same function, $f(a) = g(a)$. This being true for all $a \in A$, f and g are one and the same function from A to C^B.

To prove surjectivity, let $h \in C^{B\times A}$. Let $h_A : A \to C^B$ be defined as follows. For $a \in A$, $h_A(a)$ is the function from B to C specified by

$$h_A(a)(b) = h(\langle b,a\rangle) \qquad \text{for all} \quad b \in B$$

It can now be verified that $F(h_A) = h$. Thus every $h \in C^{B \times A}$ belongs to Im F, and F is surjective onto $C^{B \times A}$. Since it was also shown to be injective, it must be bijective to $C^{B \times A}$, establishing the equipotence of its domain $(C^B)^A$ and the codomain $C^{B \times A}$. □

EXERCISES

1. For any sets A, B, C verify that
   (a) $A \subseteq B$ is equivalent to $A \cup B = B$,
   (b) $A \times (B \cup C) = (A \times B) \cup (A \times C)$,
   (c) $A \times \emptyset = \emptyset$,
   (d) if I is a singleton, then $A \times I \simeq A$,
   (e) $A \times B \simeq B \times A$,
   (f) $A \times (B \times C) \simeq (A \times B) \times C$,
   (g) $(A \times B)^C \simeq (A^C) \times (B^C)$,
   (h) $A^{\emptyset}$ is a singleton,
   (i) $\emptyset^A = \emptyset$ unless $A = \emptyset$,
   (j) if I is a singleton, then $A^I \simeq A$ and I^A is a singleton,
   (k) if $A \neq \emptyset$, then the set of constant functions $A \to B$ is equipotent to B,
   (l) for $A \neq \emptyset$, a function $f : A \to B$ is injective if and only if there is a function $g : B \to A$ with $g \circ f = id_A$.
2. Write and run a computer program that produces a complete list of members of the set $\mathcal{PPPPPP}(\emptyset)$. Introduce any notation you wish to make the printout readable.

2. CARDINAL AND ORDINAL NUMBERS

An *ordinal* (or *ordinal number*) is a set α satisfying the following conditions:

(i) every element of α is a subset of α, $\alpha \subseteq \mathcal{P}(\alpha)$,
(ii) for $b, c \in \alpha$, $c \subset b$ if and only if $c \in b$,

(iii) every nonvoid subset S of α, $\emptyset \subset S \subseteq \alpha$, has a member $p \in S$ that is a subset of all members $s \in S$, $p \subseteq s$; p is then called the *first element* of S.

Examples. The null set $\emptyset$, the singleton $\{\emptyset\}$, and the pair $\{\emptyset, \{\emptyset\}\}$ are ordinals. There will be many more.

Condition (i) tells us that for an ordinal α, if $b \in \alpha$ and $c \in b$, then $c \in \alpha$. The reader can see that every element b of an ordinal α is again an ordinal.

Condition (ii) implies that an ordinal never belongs to itself, because no set can be a proper subset of itself. Similarly, an ordinal never belongs to any of its own elements.

From these facts it easily follows that for every ordinal α, the set $\alpha' = \alpha \cup \{\alpha\}$ is again an ordinal, called the *successor* of α. This allows us to construct some very important ordinals. We define:

$0 = \emptyset$	"*zero*"
$1 = 0' = \{\emptyset\}$	"*one*"
$2 = 1' = \{\emptyset, \{\emptyset\}\}$	"*two*"
$3 = 2'$	"*three*"
$4 = 3'$	"*four*"
$5 = 4'$	"*five*"
$6 = 5'$	"*six*"
$7 = 6'$	"*seven*"
$8 = 7'$	"*eight*"
$9 = 8'$	"*nine*"

Proposition 6 *For any ordinals α and β, we have $\alpha \subset \beta$ if and only if $\alpha \in \beta$.*

Proof. If $\alpha \in \beta$, then $\alpha \subseteq \beta$ by the definition of an ordinal applied to β. Also $\alpha \neq \beta$ because $\beta \notin \beta$. Thus $\alpha \subset \beta$.

If $\alpha \subset \beta$, then let φ be the first element of

$$S = \{x \in \beta : x \notin \alpha\}$$

For any $a \in \alpha$, the first element of $\{a, \varphi\}$ cannot be φ, for that would imply $\varphi \in \alpha$, so it is a, and thus $a \in \varphi$. Hence $\alpha \subseteq \varphi$. If we had the

strict inclusion $\alpha \subset \varphi$, then φ being a subset of the ordinal β, it would have as a member some element x of S, $x \in \varphi$. But by definition of S and φ we would also have $\varphi \in x$, which is impossible because an ordinal never belongs to any of it own elements. Thus $\alpha = \varphi$ and therefore $\alpha \in \beta$, which proves the proposition. □

In view of this proposition, instead of writing $\alpha \subset \beta$ or $\alpha \in \beta$ when such is the case, it is customary to write $\alpha < \beta$ and to say that α is *less than* β, or that β is *greater than* α. We write $\alpha \leq \beta$ to mean that "$\alpha < \beta$ or $\alpha = \beta$" and we say that α is *less than or equal to* β, or β is *greater than or equal to* α. For example, every ordinal α is less than its successor α', but neither $\alpha' < \alpha$ nor $\alpha' \leq \alpha$ is true. The inequality $\alpha \leq \beta$ is equivalent to $\alpha \subseteq \beta$.

A key property of the relation $\leq$ is that it permits the comparison of any two ordinals. For assume that neither $\alpha \leq \beta$ nor $\beta \leq \alpha$ holds. The set σ of common elements of α and β is easily seen to be an ordinal. By assumption, σ is distinct both from α and β. But then, by Proposition 6, σ belongs to both α and β, i.e., σ belongs to σ, which is impossible. This proves that at least one of $\alpha \leq \beta$ or $\beta \leq \alpha$ must hold. Combining with earlier remarks, we obtain:

Proposition 7 *Let α, β, and γ be ordinals.*

(i) $\alpha \leq \alpha$	(*reflexivity*)
(ii) *if* $\alpha \leq \beta$ *and* $\beta \leq \alpha$, *then* $\alpha = \beta$	(*antisymmetry*)
(iii) *if* $\alpha \leq \beta$ *and* $\beta \leq \gamma$, *then* $\alpha \leq \gamma$	(*transitivity*)
(iv) *at least one of* $\alpha \leq \beta$ *or* $\beta \leq \alpha$ *holds*	(*total comparability*)

Indeed a property stronger than total comparability holds:

Proposition 8 *Every nonempty set S of ordinals has an element φ that is less than any other element of S.*

Proof. Take any $\alpha \in S$. Let S_α be the set of those elements of S that are less than α. If $S_\alpha = \emptyset$, then let $\varphi = \alpha$. Otherwise S_α is a nonvoid subset of α, and we let φ be the first element of S_α. □

Corollary. *The union of any set of ordinals is an ordinal.*

The ordinal $\varphi \in S$ whose existence was established by Proposition 8 is called the *first element* of S, which is in accordance with the earlier use of this term.

Every ordinal number α is a set, namely the set of ordinals less than α. The union $\cup\alpha$ is always an ordinal and $\cup\alpha \leq \alpha$. Now, either $\cup\alpha < \alpha$ or $\cup\alpha = \alpha$. If $\cup\alpha < \alpha$, then let ρ be an element of α that does not belong to $\cup\alpha$. We have $\cup\alpha \leq \rho$. But since $\rho \in \alpha$, also $\rho \subseteq \cup\alpha$. Thus $\rho = \cup\alpha$, and the only element of α not in $\cup\alpha$ is $\rho = \cup\alpha$, i.e.,

$$\alpha = (\cup\alpha) \cup \{\cup\alpha\}$$

which means that α is the successor of $\cup\alpha$. Can α be at the same time the successor of some other ordinal β? The answer is no, because from $\alpha = \beta \cup \{\beta\}$ it follows that

$$\cup\alpha = (\cup\beta) \cup \beta = \beta$$

This argument also shows that α is not the successor of any ordinal β if $\cup\alpha = \alpha$. There are two kinds of nonzero ordinals α. On the one hand, there are those for which $\cup\alpha < \alpha$. Then α is the successor of $\cup\alpha$, and $\cup\alpha$ is termed the *predecessor* of α. On the other hand, there are those ordinals α for which $\cup\alpha = \alpha$. These have no predecessor, and they are called *limit ordinals*. Every limit ordinal is the union of lesser ordinals. An ordinal α that is either 0 or such that $\cup\alpha < \alpha$ is said to be of the *first kind*, while nonzero limit ordinals are sometimes said to be of the *second kind*. We now arrive at a most important concept: an ordinal is called *finite* if it is of the first kind and all its elements are also of the first kind. An ordinal that is not finite is called *infinite*. Finite ordinals are also called *natural numbers*.

Examples. The ordinals $0, 1, 2, \ldots, 9$ defined earlier are natural numbers. The successor of any natural number is again a natural number.

But does there exist any infinite ordinal? We are unable to prove it. We have seen how to make "one" out of "zero," "two" out of "one," and so on to trillions. However, even a megaquadrillion is just finite dust. It is time to postulate three new axioms.

(A6) Axiom of Existential Infinity. *There is a set to which all finite ordinals belong.*

(A7) Axiom of Limited Infinity. *There is no set having among its members sets equipotent to every ordinal.*

Otherwise stated, for every set S, there is an ordinal α such that no set belonging to S is equipotent to α.

(A8) Axiom of Choice. *For every set of nonvoid sets S, there exists a function $c : S \to \cup S$ such that $c(A) \in A$ for every $A \in S$.*

The function c is called a *choice function*; for each $A \in S$ it is said to *choose* the element $c(A)$ in the set A.

An immediate consequence of the Axiom of Existential Infinity is that there is a set whose elements are precisely the natural numbers. In view of the paramount importance that it claims in the spiritual life of mathematicians, *the set of all natural numbers* is denoted by the Greek letter ω. The Axiom of Limited Infinity, on the other hand, implies that the "set of all ordinals" is nonexistent.

With the intention of using ordinal numbers for enumeration, we now further develop the theory of equipotence, in particular as regarding ordinals. For any sets S and R the *difference* $S \setminus R$ is defined by

$$S \setminus R = \{x \in S : x \notin R\}$$

Counting Lemma. *If a set S is equipotent to a set T, and if s and t are elements of S and T, respectively, then $S \setminus \{s\}$ and $T \setminus \{t\}$ are equipotent.*

Proof. If $f : S \to T$ is a bijection, and if $f(s) = t$, then

$$g = \{\langle x, y\rangle \in f : x \neq s,\ y \neq t\}$$

is a bijection from $S \setminus \{s\}$ to $T \setminus \{t\}$. If $f(s) \neq t$, then let $r \in S$ such that $f(r) = t$. Now

$$g = \{\langle x, y\rangle \in f : x \neq s,\ y \neq t\} \cup \{\langle r, f(s)\rangle\}$$

is a bijection from $S \setminus \{s\}$ to $T \setminus \{t\}$. □

It is now easy to show that no natural number is equipotent to any other natural number. (If this is not true, let n be the first natural number equipotent to some natural number distinct from itself, say to $m \neq n$. As no bijection can exist between the empty set and a

nonempty set, neither of n or m is 0. Then by the Counting Lemma, the predecessors of n and m are equipotent, contradicting the definition of n.) Thus a natural number is not equipotent to any of its own elements. This is not true for ordinal numbers in general. For example, the successor of ω, the ordinal $\omega' = \omega \cup \{\omega\}$, is equipotent to ω. A bijection $f : \omega' \to \omega$ can be defined by

$$f(n) = n' \qquad \text{for all} \quad n \in \omega', \quad n \neq \omega, \quad \text{and} \quad f(\omega) = 0$$

This observation motivates the following definition. A *cardinal* (or *cardinal number*) is an ordinal that is not equipotent to any of its own elements (i.e., not equipotent to any ordinal less than itself). Thus natural numbers are cardinal numbers. Let us verify that so is ω. Were this not so, there would be a smallest $n \in \omega$ equipotent to ω. Obviously $n \neq 0$, so let m be the predecessor of n. By the Counting Lemma, the sets

$$m = n \setminus \{m\} \qquad \text{and} \qquad \overline{\omega} = \omega \setminus \{0\}$$

would be equipotent. But $\overline{\omega}$ is also equipotent to ω, via the bijection

$$f = \{\langle n, n' \rangle : n \in \omega\}$$

so by transitivity (Proposition 4) m would be equipotent to ω, contradicting the minimal choice of n.

Zermelo's Theorem (First Formulation). *Every set is equipotent to a cardinal.*

Proof. Observe first that it will be enough to prove that every set S is equipotent to some ordinal α. For if α is not a cardinal, then let β be the first element of α that is equipotent to α. Obviously β is a cardinal equipotent to S.

To prove that every set S is equipotent to an ordinal, we use the Axiom of Choice, which assures us of the existence of a choice function c from the set $\mathcal{P}^*(S)$ of nonempty subsets of S into $\cup\mathcal{P}^*(S) = S$. With a fixed choice function c in mind, we call *ordinal function* into S any injection $f : \alpha \to S$ from some ordinal α into S, such that for every $\beta \in \alpha$

$$f(\beta) = c(S \setminus f[\beta])$$

In particular, if $\alpha \neq 0$, then $f(0) = c(S)$ for every ordinal function $f : \alpha \to S$. Further, we claim that if $f : \alpha \to S$ and $g : \rho \to S$ are ordinal

functions and $\alpha \leq \rho$, then $f(\beta) = g(\beta)$ for every $\beta \in \alpha$. For if β were the first element of α for which $f(\beta) \neq g(\beta)$, then $f(\gamma) = g(\gamma)$ for all $\gamma \in \beta$ and the sets

$$S_f = S \setminus f[\beta]$$
$$S_g = S \setminus g[\beta]$$

would be the same, hence $c(S_f) = c(S_g)$. But $f(\beta) = c(S_f)$ and $g(\beta) = c(S_g)$, and then $f(\beta) = g(\beta)$, proving our claim. This implies in particular that, with respect to a fixed choice function c, there can be at most one ordinal function $\alpha \to S$ for any ordinal α.

Ordinal functions are injective. Therefore, their images are equipotent to their ordinal domains. Hence, by the Axiom of Limited Infinity, there is an ordinal σ without ordinal function $f : \sigma \to S$. Then define the ordinal β as follows. If for every ordinal $\rho < \sigma$ there are ordinal functions from ρ to S, then let $\beta = \sigma$. Otherwise let β be the first element of ρ for which no ordinal function exists from β to S. In either case, β has the following properties: there is no ordinal function $\beta \to S$; and for every $\rho < \beta$ there is an ordinal function $\rho \to S$. Observe that β cannot be a limit ordinal, for in that case we could define an ordinal function $\beta \to S$ as the union of all ordinal functions with domains less than β. Also, β cannot be 0 for the empty set is surely an ordinal function $0 \to S$. Thus β has a predecessor α, and there is an ordinal function f from α to S. Is f surjective onto S? If it were not, then letting

$$a = c(S \setminus \operatorname{Im} f)$$

we could define an ordinal function g by $g = f \cup \{\langle \alpha, a \rangle\}$ on the domain β, which is impossible. Thus f must be surjective, and since ordinal functions are injective, this implies that f is a bijection from α to S. The proof is finished. □

Since by definition no cardinal is equipotent to any other cardinal, it follows from Zermelo's Theorem that every set S is equipotent to a unique cardinal, called the *cardinal* (or *cardinality*) of S and denoted by $\operatorname{Card} S$. We also say that $\operatorname{Card} S$ is the *number of elements* of S. Note that $A \subseteq B$ implies $\operatorname{Card} S \leq \operatorname{Card} B$. Sets are called *finite* or *infinite* according to whether their cardinal is finite or infinite. The following is now elementary:

Proposition 9 *Two sets are equipotent if and only if they have the same cardinal.*

As, by an earlier remark, there is no injection, and therefore no bijection, from $\mathcal{P}(\omega)$ to ω, $\mathrm{Card}\,\mathcal{P}(\omega)$ is distinct from ω. Thus there are infinite cardinals other than ω.

We now introduce a terminological redundancy. A *family* is simply a function $f : A \to B$. The domain A is called the *index set* of the family, and f is said to be a *family (of elements of B) indexed by (the elements of) A*. For $i \in A$, the element $f(i)$ of B is denoted by b_i, while the family f itself is often denoted by $(b_i : i \in A)$ or $(b_i)_{i \in A}$. For example, for any set S, the identity function id_S is nothing else but the family $(x : x \in S)$. The image set of a family $(b_i : i \in A)$ is denoted by $\{b_i : i \in A\}$. Families generalize the set concept. For example, we define the *union of a family of sets* $(b_i)_{i \in A}$, in symbols $\bigcup_{i \in A} b_i$, as the set

$$\cup\{b \in B : b = b_i \text{ for some } i \in A\}$$

It is important to notice that in a family $(b_i)_{i \in A}$ we may have $b_i = b_j$ even if the indices i and j are distinct. A family indexed by an ordinal is usually called a *sequence*. A sequence indexed by a natural number n is called an *n-tuple* (*couple, triple, quadruple, quintuple* for $n = 2,3,4,5$) and it is usually written as a string of n elements of B, possibly in brackets and separated by commas, such as (u), $(u\,v)$, $(u\,v\,w)$, $(u\,v\,w\,t)$ for $n = 1,2,3,4$ and

$$(u_0, u_1, \ldots, u_i, \ldots)$$

in general. The position i of u_i, $0 \leq i < n$, in the string indicates that the sequence in question, as a function from n to B, maps i to u_i. The image set of the sequence can be written accordingly as

$$\{u_0, u_1, \ldots, u_i, \ldots\}$$

or explicitly as $\{u, v, w\}$, $\{u, v, w, t\}$ for $n = 3,4$, etc.

Let $(A_i)_{i \in I}$ be a family of sets indexed by a set I. The *product $\prod_{i \in I} A_i$ of the family* is the set of all functions $f : I \to \bigcup_{i \in I} A_i$ such that

$$f(i) \in A_i \qquad \text{for every} \quad i \in I$$

For each $j \in I$ the function $\mathrm{pr}_j : \prod A_i \to A_j$ defined on the product set by $\mathrm{pr}_j(f) = f(j)$ is called the jth *projection*.