CHAPTER ONE

Sets and Functions

It is sometimes said that mathematics *is* the study of sets and functions. Naturally, this oversimplifies matters; but it does come as close to the truth as an aphorism can.

The study of sets and functions leads two ways. One path goes down, into the abysses of logic, philosophy, and the foundations of mathematics. The other goes up, onto the highlands of mathematics itself, where these concepts are indispensable in almost all of pure mathematics as it is today. Needless to say, we follow the latter course. We regard sets and functions as tools of thought, and our purpose in this chapter is to develop these tools to the point where they are sufficiently powerful to serve our needs through the rest of this book.

As the reader proceeds, he will come to understand that the words *set* and *function* are not as simple as they may seem. In a sense, they are simple; but they are potent words, and the quality of simplicity they possess is that which lies on the far side of complexity. They are like seeds, which are primitive in appearance but have the capacity for vast and intricate development.

1. SETS AND SET INCLUSION

We adopt a naive point of view in our discussion of sets and assume that the concepts of an element and of a set of elements are intuitively clear. By an *element* we mean an object or entity of some sort, as, for example, a positive integer, a point on the real line (= a real number),

or a point in the complex plane (= a complex number). A *set* is a collection or aggregate of such elements, considered together or as a whole. Some examples are furnished by the set of all even positive integers, the set of all rational points on the real line, and the set of all points in the complex plane whose distance from the origin is 1 (= the unit circle in the plane). We reserve the word *class* to refer to a set of sets. We might speak, for instance, of the class of all circles in a plane (thinking of each circle as a set of points). It will be useful in the work we do if we carry this hierarchy one step further and use the term *family* for a set of classes. One more remark: the words element, set, class, and family are not intended to be rigidly fixed in their usage; we use them fluidly, to express varying attitudes toward the mathematical objects and systems we study. It is entirely reasonable, for instance, to think of a circle not as a set of points, but as a single entity in itself, in which case we might justifiably speak of the set of all circles in a plane.

There are two standard notations available for designating a particular set. Whenever it is feasible to do so, we can list its elements between braces. Thus $\{1, 2, 3\}$ signifies the set consisting of the first three positive integers, $\{1, i, -1, -i\}$ is the set of the four fourth roots of unity, and $\{\pm 1, \pm 3, \pm 5, \ldots\}$ is the set of all odd integers. This manner of specifying a set, by listing its elements, is unworkable in many circumstances. We are then obliged to fall back on the second method, which is to use a property or attribute that characterizes the elements of the set in question. If P denotes a certain property of elements, then $\{x:P\}$ stands for the set of all elements x for which the property P is meaningful and true. For example, the expression

$$\{x:x \text{ is real and irrational}\},$$

which we read *the set of all x such that x is real and irrational*, denotes the set of all real numbers which cannot be written as the quotient of two integers. The set under discussion contains all those elements (and no others) which possess the stated property. The three sets of numbers described at the beginning of this paragraph can be written either way:

$$\{1, 2, 3\} = \{n:n \text{ is an integer and } 0 < n < 4\},$$
$$\{1, i, -1, -i\} = \{z:z \text{ is a complex number and } z^4 = 1\},$$
and
$$\{\pm 1, \pm 3, \pm 5, \ldots\} = \{n:n \text{ is an odd integer}\}.$$

We often shorten our notation. For instance, the last two sets mentioned might perfectly well be written $\{z:z^4 = 1\}$ and $\{n:n \text{ is odd}\}$. Our purpose is to be clear and to avoid misunderstandings, and if this can be achieved with less notation, so much the better. In the same vein we can

write

$$\text{the } \textit{unit circle} = \{z:|z| = 1\},$$
$$\text{the } \textit{closed unit disc} = \{z:|z| \leq 1\},$$

and

$$\text{the } \textit{open unit disc} = \{z:|z| < 1\}.$$

We use a special system of notation for designating intervals of various kinds on the real line. If a and b are real numbers such that $a < b$, then the following symbols on the left are defined to be the indicated sets on the right:

$$[a,b] = \{x:a \leq x \leq b\},$$
$$(a,b] = \{x:a < x \leq b\},$$
$$[a,b) = \{x:a \leq x < b\},$$
$$(a,b) = \{x:a < x < b\}.$$

We speak of these as the closed, the open-closed, the closed-open, and the open intervals from a to b. In particular, [0,1] is the *closed unit interval*, and (0,1) is the *open unit interval*.

There are certain logical difficulties which arise in the foundations of the theory of sets (see Problem 1). We avoid these difficulties by assuming that each discussion in which a number of sets are involved takes place in the context of a single fixed set. This set is called the *universal set*. It is denoted by U in this section and the next, and every set mentioned is assumed to consist of elements in U. In later chapters there will always be on hand a given space within which we work, and this will serve without further comment as our universal set.[1] It is often convenient to have available in U a set containing no elements whatever; we call this the *empty set* and denote it by the symbol $\emptyset$. A set is said to be *finite* if it is empty or consists of n elements for some positive integer n; otherwise, it is said to be *infinite*.

We usually denote elements by small letters and sets by large letters. If x is an element and A is a set, the statement that x is an element of A (or belongs to A, or is contained in A) is symbolized by $x \in A$. We denote the negation of this, namely, the statement that x is not an element of A, by $x \notin A$.

Two sets A and B are said to be *equal* if they consist of exactly the same elements; we denote this relation by $A = B$ and its negation by $A \neq B$. We say that A is a *subset* of B (or is contained in B) if each element of A is also an element of B. This relation is symbolized by $A \subseteq B$. We sometimes express this by saying that B is a *superset* of A (or con-

[1] The words *set* and *space* are often used in loose contrast to one another. A set is merely an amorphous collection of elements, without coherence or form. When some kind of algebraic or geometric structure is imposed on a set, so that its elements are organized into a systematic whole, then it becomes a space.

tains A). $A \subseteq B$ allows for the possibility that A and B might be equal. If A is a subset of B and is not equal to B, we say that A is a *proper subset* of B (or is properly contained in B). This relation is denoted by $A \subset B$. We can also express $A \subset B$ by saying that B is a *proper superset* of A (or properly contains A). The relation $\subseteq$ is usually called *set inclusion.*

We sometimes reverse the symbols introduced in the previous paragraph. Thus $A \subseteq B$ and $A \subset B$ are occasionally written in the equivalent forms $B \supseteq A$ and $B \supset A$.

It will often be convenient to have a symbol for logical implication, and $\Rightarrow$ is the symbol we use. If p and q are statements, then $p \Rightarrow q$ means that p *implies* q, or that if p is true, then q is also true. Similarly, $\Leftrightarrow$ is our symbol for two-way implication or logical equivalence. It means that the statement on each side implies the statement on the other, and is usually read *if and only if*, or *is equivalent to.*

The main properties of set inclusion are obvious. They are the following:

(1) $A \subseteq A$ for every A;

(2) $A \subseteq B$ and $B \subseteq A \Rightarrow A = B$;

(3) $A \subseteq B$ and $B \subseteq C \Rightarrow A \subseteq C$.

It is quite important to observe that (1) and (2) can be combined into the single statement that $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$. This remark contains a useful principle of proof, namely, that the only way to show that two sets are equal, apart from merely inspecting them, is to show that each is a subset of the other.

Problems

1. Perhaps the most famous of the logical difficulties referred to in the text is *Russell's paradox.* To explain what this is, we begin by observing that a set can easily have elements which are themselves sets, e.g., $\{1, \{2,3\}, 4\}$. This raises the possibility that a set might well contain itself as one of its elements. We call such a set an *abnormal* set, and any set which does not contain itself as an element we call a *normal* set. Most sets are normal, and if we suspect that abnormal sets are in some way undesirable, we might try to confine our attention to the set N of all normal sets. Someone is now sure to ask, Is N itself normal or abnormal? It is evidently one or the other, and it cannot be both. Show that if N is normal, then it must be abnormal. Show also that if N is abnormal, then it must be normal. We see in this way that each of our two alternatives is self-contradictory, and it seems to be the assumption that N exists as a set which has brought us to this impasse. For further discussion of these matters, we refer the interested reader to Wilder [42, p. 55]

or Fraenkel and Bar-Hillel [10, p. 6]. Russell's own account of the discovery of his paradox can be found in Russell [36, p. 75].

2. The symbol we have used for set inclusion is similar to that used for the familiar order relation on the real line: if x and y are real numbers, $x \leq y$ means that $y - x$ is non-negative. The order relation on the real line has all the properties mentioned in the text:

(1′) $x \leq x$ for every x;

(2′) $x \leq y$ and $y \leq x \Rightarrow x = y$;

(3′) $x \leq y$ and $y \leq z \Rightarrow x \leq z$.

It also has an important additional property:

(4′) for any x and y, either $x \leq y$ or $y \leq x$.

Property (4′) says that any two real numbers are comparable with respect to the relation in question, and it leads us to call the order relation on the real line a *total* (or *linear*) *order relation.* Show by an example that this property is not possessed by set inclusion. It is for this reason that set inclusion is called a *partial order relation.*

3. (*a*) Let U be the single-element set $\{1\}$. There are two subsets, the empty set $\emptyset$ and $\{1\}$ itself. If A and B are arbitrary subsets of U, there are four possible relations of the form $A \subseteq B$. Count the number of true relations among these.

(*b*) Let U be the set $\{1,2\}$. There are four subsets. List them. If A and B are arbitrary subsets of U, there are 16 possible relations of the form $A \subseteq B$. Count the number of true ones.

(*c*) Let U be the set $\{1, 2, 3\}$. There are 8 subsets. What are they? There are 64 possible relations of the form $A \subseteq B$. Count the number of true ones.

(*d*) Let U be the set $\{1, 2, \ldots, n\}$ for an arbitrary positive integer n. How many subsets are there? How many possible relations of the form $A \subseteq B$ are there? Can you make an informed guess as to how many of these are true?

2. THE ALGEBRA OF SETS

In this section we consider several useful ways in which sets can be combined with one another, and we develop the chief properties of these operations of combination.

As we emphasized above, all the sets we mention in this section are assumed to be subsets of our universal set U. U is the *frame of reference,* or the *universe,* for our present discussions. In our later work the frame of reference in a particular context will naturally depend on what ideas we happen to be considering. If we find ourselves studying sets of real

numbers, then U is the set R of all real numbers. If we wish to study sets of complex numbers, then we take U to be the set C of all complex numbers. We sometimes want to narrow the frame of reference and to consider (for instance) only subsets of the closed unit interval [0,1], or of the closed unit disc $\{z:|z| \leq 1\}$, and in these cases we choose U accordingly. Generally speaking, the universal set U is at our disposal, and we are free to select it to fit the needs of the moment. For the present, however, U is to be regarded as a fixed but arbitrary set. This generality allows us to apply the ideas we develop below to any situation which arises in our later work.

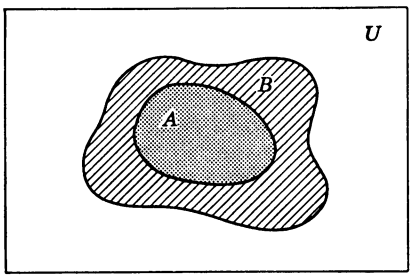

Fig. 1. Set inclusion.

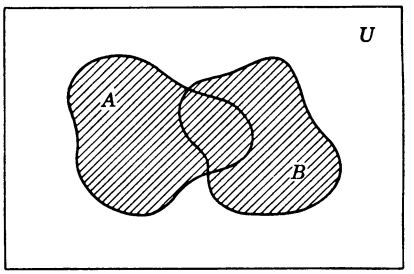

Fig. 2. The union of A and B.

It is extremely helpful to the imagination to have a geometric picture available in terms of which we can visualize sets and operations on sets. A convenient way to accomplish this is to represent U by a rectangular area in a plane, and the elements which make up U by the points of this area. Sets can then be pictured by areas within this rectangle, and diagrams can be drawn which illustrate operations on sets and relations between them. For instance, if A and B are sets, then Fig. 1 represents the circumstance that A is a subset of B (we think of each set as consisting of all points within the corresponding closed curve). Diagrammatic thought of this kind is admittedly loose and imprecise; nevertheless, the reader will find it invaluable. No mathematics, however abstract it may appear, is ever carried on without the help of mental images of some kind, and these are often nebulous, personal, and difficult to describe.

The first operation we discuss in the algebra of sets is that of forming unions. The *union* of two sets A and B, written $A \cup B$, is defined to be the set of all elements which are in either A or B (including those which are in both). $A \cup B$ is formed by lumping together the elements of A and those of B and regarding them as constituting a single set. In Fig. 2, $A \cup B$ is indicated by the shaded area. The above

definition can also be expressed symbolically:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

The operation of forming unions is commutative and associative:

$$A \cup B = B \cup A \qquad \text{and} \qquad A \cup (B \cup C) = (A \cup B) \cup C.$$

It has the following additional properties:

$$A \cup A = A,\ A \cup \emptyset = A, \text{ and } A \cup U = U.$$

We also note that

$$A \subseteq B \Leftrightarrow A \cup B = B,$$

so set inclusion can be expressed in terms of this operation.

Our next operation is that of forming intersections. The *intersection* of two sets A and B, written $A \cap B$, is the set of all elements which are in both A and B. In symbols,

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

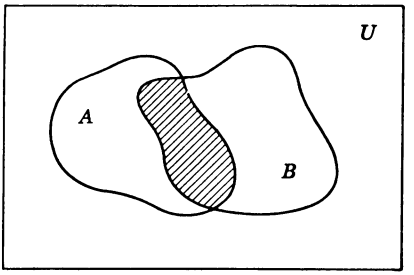

Fig. 3. The intersection of A and B.

$A \cap B$ is the common part of the sets A and B. In Fig. 3, $A \cap B$ is represented by the shaded area. If $A \cap B$ is non-empty, we express this by saying that A *intersects* B. If, on the other hand, it happens that A and B have no common part, or equivalently that $A \cap B = \emptyset$, then we say that A *does not intersect* B, or that A and B are *disjoint;* and a class of sets in which all pairs of distinct sets are disjoint is called a *disjoint class* of sets. The operation of forming intersections is also commutative and associative:

$$A \cap B = B \cap A \qquad \text{and} \qquad A \cap (B \cap C) = (A \cap B) \cap C.$$

It has the further properties that

$$A \cap A = A,\ A \cap \emptyset = \emptyset, \text{ and } A \cap U = A;$$

and since

$$A \subseteq B \Leftrightarrow A \cap B = A,$$

we see that set inclusion can also be expressed in terms of forming intersections.

We have now defined two of the fundamental operations on sets, and we have seen how each is related to set inclusion. The next obvious step is to see how they are related to one another. The facts here are given by

the *distributive laws:*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

These properties depend only on simple logic applied to the meanings of

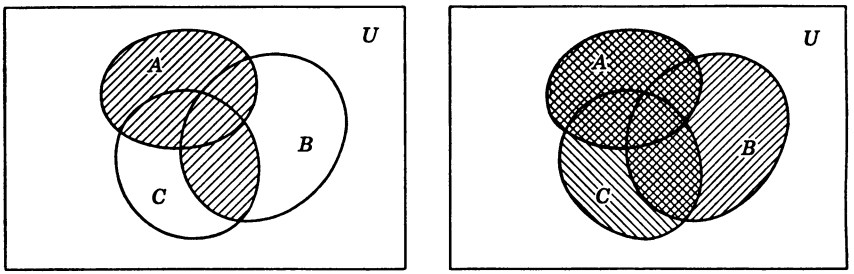

Fig. 4. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

the symbols involved. For instance, the first of the two distributive laws says that an element is in A and is in B or C precisely when it is in A and B or is in A and C. We can convince ourselves intuitively of the validity of these laws by drawing pictures. The second distributive law is illustrated in Fig. 4, where $A \cup (B \cap C)$ is formed on the left by shading and $(A \cup B) \cap (A \cup C)$ on the right by cross-shading. A moment's consideration of these diagrams ought to convince the reader that one obtains the same set in each case.

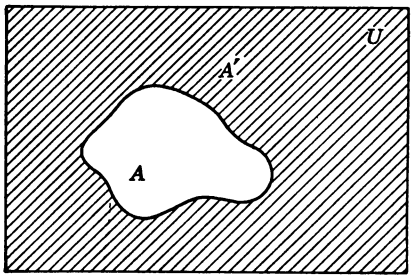

Fig. 5. The complement of A.

The last of our major operations on sets is the formation of complements. The *complement* of a set A, denoted by A', is the set of all elements which are not in A. Since the only elements we consider are those which make up U, it goes without saying—but it ought to be said —that A' consists of all those elements in U which are not in A. Symbolically,

$$A' = \{x : x \notin A\}.$$

Figure 5 (in which A' is shaded) illustrates this operation. The operation of forming complements has the following obvious properties:

$$(A')' = A,\ \emptyset' = U,\ U' = \emptyset,$$
$$A \cup A' = U, \text{ and } A \cap A' = \emptyset.$$

Further, it is related to set inclusion by

$$A \subseteq B \Leftrightarrow B' \subseteq A'$$

and to the formation of unions and intersections by

$$(A \cup B)' = A' \cap B' \qquad \text{and} \qquad (A \cap B)' = A' \cup B'. \tag{1}$$

The first equation of (1) says that an element is not in either of two sets precisely when it is outside of both, and the second says that it is not in both precisely when it is outside of one or the other.

The operations of forming unions and intersections are primarily binary operations; that is, each is a process which applies to a pair of sets and yields a third. We have emphasized this by our use of parentheses to indicate the order in which the operations are to be performed, as in $(A_1 \cup A_2) \cup A_3$, where the parentheses direct us first to unite A_1 and A_2, and then to unite the result of this with A_3. Associativity makes it possible to dispense with parentheses in an expression like this and to write $A_1 \cup A_2 \cup A_3$, where we understand that these sets are to be united in any order and that the order in which the operations are performed is irrelevant. Similar remarks apply to $A_1 \cap A_2 \cap A_3$. Furthermore, if $\{A_1, A_2, \ldots, A_n\}$ is any finite class of sets, then we can form

$$A_1 \cup A_2 \cup \cdots \cup A_n \qquad \text{and} \qquad A_1 \cap A_2 \cap \cdots \cap A_n$$

in much the same way without any ambiguity of meaning whatever. In order to shorten the notation, we let $I = \{1, 2, \ldots, n\}$ be the set of subscripts which index the sets under consideration. I is called the *index set.* We then compress the symbols for the union and intersection just mentioned to $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$. As long as it is quite clear what the index set is, we can write this union and intersection even more briefly, in the form $\bigcup_i A_i$ and $\bigcap_i A_i$. For the sake of both brevity and clarity, these sets are often written $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i$.

These extensions of our ideas and notations don't reach nearly far enough. It is often necessary to form unions and intersections of large (really large!) classes of sets. Let $\{A_i\}$ be an entirely arbitrary class of sets indexed by a set I of subscripts. Then

$$\bigcup_{i \in I} A_i = \{x : x \in A_i \text{ for at least one } i \in I\}$$

and

$$\bigcap_{i \in I} A_i = \{x : x \in A_i \text{ for every } i \in I\}$$

define their *union* and *intersection.* As above, we usually abbreviate these notations to $\bigcup_i A_i$ and $\bigcap_i A_i$; and if the class $\{A_i\}$ consists of a sequence of sets, that is, if $\{A_i\} = \{A_1, A_2, A_3, \ldots\}$, then their union and intersection are often written in the form $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$. Observe that we did not require the class $\{A_i\}$ to be non-empty. If it does

happen that this class is empty, then the above definitions give (remembering that all sets are subsets of U) $\cup_i A_i = \emptyset$ and $\cap_i A_i = U$. The second of these facts amounts to the following statement: if we require of an element that it belong to each set in a given class, and if there are no sets present in the class, then every element satisfies this requirement. If we had not made the agreement that the only elements under consideration are those in U, we would not have been able to assign a meaning to the intersection of an empty class of sets. A moment's consideration makes it clear that Eqs. (1) are valid for arbitrary unions and intersections:

$$(\cup_i A_i)' = \cap_i A_i' \qquad \text{and} \qquad (\cap_i A_i)' = \cup_i A_i'. \tag{2}$$

It is instructive to verify these equations for the case in which the class $\{A_i\}$ is empty.

We conclude our treatment of the general theory of sets with a brief discussion of certain special classes of sets which are of considerable importance in topology, logic, and measure theory. We usually denote classes of sets by capital letters in boldface.

First, some general remarks which will be useful both now and later, especially in connection with topological spaces. We shall often have occasion to speak of *finite unions* and *finite intersections*, by which we mean unions and intersections of finite classes of sets, and by a finite class of sets we always mean one which is empty or consists of n sets for some positive integer n. If we say that a class $\mathbf{A}$ of sets is closed under the formation of finite unions, we mean that $\mathbf{A}$ contains the union of each of its finite subclasses; and since the empty subclass qualifies as a finite subclass of $\mathbf{A}$, we see that its union, the empty set, is necessarily an element of $\mathbf{A}$. In the same way, a class of sets which is closed under the formation of finite intersections necessarily contains the universal set.

Now for the special classes of sets mentioned above. For the remainder of this section we specifically assume that the universal set U is non-empty. A *Boolean algebra of sets* is a non-empty class $\mathbf{A}$ of subsets of U which has the following properties:

(1) A and $B \in \mathbf{A} \Rightarrow A \cup B \in \mathbf{A}$;

(2) A and $B \in \mathbf{A} \Rightarrow A \cap B \in \mathbf{A}$;

(3) $A \in \mathbf{A} \Rightarrow A' \in \mathbf{A}$.

Since $\mathbf{A}$ is assumed to be non-empty, it must contain at least one set A. Property (3) shows that A' is in $\mathbf{A}$ along with A, and since $A \cap A' = \emptyset$ and $A \cup A' = U$, (1) and (2) guarantee that $\mathbf{A}$ contains the empty set and the universal set. Since the class consisting only of the empty set and the universal set is clearly a Boolean algebra of sets, these two distinct sets are the only ones which every Boolean algebra of sets must

contain. It is equally clear that the class of all subsets of U is also a Boolean algebra of sets. There are many other less trivial kinds, and their applications are manifold in fields of study as diverse as statistics and electronics.

Let **A** be a Boolean algebra of sets. It is obvious that if $\{A_1, A_2, \ldots, A_n\}$ is a non-empty finite subclass of **A**, then

$$A_1 \cup A_2 \cup \cdots \cup A_n \quad \text{and} \quad A_1 \cap A_2 \cap \cdots \cap A_n$$

are both sets in **A**; and since **A** contains the empty set and the universal set, it is easy to see that **A** is a class of sets which is closed under the formation of finite unions, finite intersections, and complements. We now go in the other direction, and let **A** be a class of sets which is closed under the formation of finite unions, finite intersections, and complements. By these assumptions, **A** automatically contains the empty set and the universal set, so it is non-empty and is easily seen to be a Boolean algebra of sets. We conclude from these remarks that Boolean algebras of sets can be described alternatively as classes of sets which are closed under the formation of finite unions, finite intersections, and complements. It should be emphasized once again that when discussing Boolean algebras of sets we always assume that the universal set is non-empty.

One final comment. We speak of Boolean algebras *of sets* because there are other kinds of Boolean algebras than those which consist of sets, and we wish to preserve the distinction. We explore this topic further in our Appendix on Boolean algebras.

Problems

1. If $\{A_i\}$ and $\{B_j\}$ are two classes of sets such that $\{A_i\} \subseteq \{B_j\}$, show that $\bigcup_i A_i \subseteq \bigcup_j B_j$ and $\bigcap_j B_j \subseteq \bigcap_i A_i$.
2. The *difference* between two sets A and B, denoted by $A - B$, is the set of all elements in A and not in B; thus $A - B = A \cap B'$. Show the following:

$$\begin{aligned} A - B = A - (A \cap B) &= (A \cup B) - B; \\ (A - B) - C &= A - (B \cup C); \\ A - (B - C) &= (A - B) \cup (A \cap C); \\ (A \cup B) - C &= (A - C) \cup (B - C); \\ A - (B \cup C) &= (A - B) \cap (A - C). \end{aligned}$$

3. The *symmetric difference* of two sets A and B, denoted by $A \Delta B$, is defined by $A \Delta B = (A - B) \cup (B - A)$; it is thus the union of

their differences in opposite orders. Show the following:

$$A \Delta (B \Delta C) = (A \Delta B) \Delta C;$$
$$A \Delta \emptyset = A; A \Delta A = \emptyset;$$
$$A \Delta B = B \Delta A;$$
$$A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C).$$

4. A *ring of sets* is a non-empty class **A** of sets such that if A and B are in **A**, then $A \Delta B$ and $A \cap B$ are also in **A**. Show that **A** must also contain the empty set, $A \cup B$, and $A - B$. Show that if a non-empty class of sets contains the union and difference of any pair of its sets, then it is a ring of sets. Show that a Boolean algebra of sets is a ring of sets.
5. Show that the class of all finite subsets (including the empty set) of an infinite set is a ring of sets but is not a Boolean algebra of sets.
6. Show that the class of all finite unions of closed-open intervals on the real line is a ring of sets but is not a Boolean algebra of sets.
7. Assuming that the universal set U is non-empty, show that Boolean algebras of sets can be described as rings of sets which contain U.

3. FUNCTIONS

Many kinds of functions occur in topology, in a great variety of situations. In our work we shall need the full power of the general concept of a function, and since its modern meaning is much broader and deeper than its elementary meaning, we discuss this concept in considerable detail and develop its main abstract properties.

Let us begin with a brief inspection of some simple examples. Consider the elementary function

$$y = x^2$$

of the real variable x. What do we have in mind when we call this a function and say that y is a function of x? In a nutshell, we are drawing attention to the fact that each real number x has linked to it a specific real number y, which can be calculated according to the rule (or law of correspondence) given by the formula. We have here a process which, applied to any real number x, does something to it (squares it) to produce another real number y (the square of x). Similarly,

$$y = x^3 - 3x \qquad \text{and} \qquad y = (x^2 + 1)^{-1}$$

are two other simple functions of the real variable x, and each is given by a rule in the form of an algebraic expression which specifies the exact manner in which the value of y depends on the value of x.

The rules for the functions we have just mentioned are expressed by formulas. In general, this is possible only for functions of a very simple kind or for those which are sufficiently important to deserve special symbols of their own. Consider, for instance, the function of the real variable x defined as follows: for each real number x, write x as an infinite decimal (using the scheme of decimal expansion in which infinite chains of 9's are avoided—in which, for example, $\frac{1}{4}$ is represented by .25000 . . . rather than by .24999 . . .); then let y be the fifty-ninth digit after the decimal point. There is of course no standard formula for this, but nevertheless it is a perfectly respectable function whose rule is given by a verbal description. On the other hand, the function $y = \sin x$ of the real variable x is so important that its rule, though fully as complicated as the one just defined, is assigned the special symbol *sin*. When discussing functions in general, we want to allow for all sorts of rules and to talk about them all at once, so we usually employ noncommittal notations like $y = f(x)$, $y = g(x)$, and so on.

Each of the functions mentioned above is defined for all real numbers x. The example $y = 1/x$ shows that this restriction is much too severe, for this function is defined only for non-zero values of x. Similarly, $y = \log x$ is defined only for positive values of x, and $y = \sin^{-1} x$ only for values of x which lie in the interval $[-1,1]$. Whatever our conception of a function may be, it should certainly be broad enough to include examples like these, which are defined only for some values of the real variable x.

In real analysis the notion of function is introduced in the following way. Let X be any non-empty set of real numbers. We say that a *function* $y = f(x)$ is defined on X if the rule f associates a definite real number y with each real number x in X. The specific nature of the rule f is totally irrelevant to the concept of a function. The set X is called the *domain* of the given function, and the set Y of all the values it assumes is called its *range*. If we speak of complex numbers here instead of real numbers, we have the notion of function as it is used in complex analysis.

This point of view toward functions is actually a bit more general than is needed for the aims of analysis, but it isn't nearly general enough for our purposes. The sets X and Y above were taken to be sets of numbers. If we now remove even this restriction and allow X and Y to be completely arbitrary non-empty sets, then we arrive at the most inclusive concept of a function. By way of illustration, suppose that X is the set of all squares in a plane and that Y is the set of all circles in the same plane. We can define a function $y = f(x)$ by requiring that the rule f associate with each square x that circle y which is inscribed in it. In general, there is no need at all for either X or Y to be a set of

numbers. All that is really necessary for a function is two non-empty sets X and Y and a rule f which is meaningful and unambiguous in assigning to each element x in X a specific element y in Y.

With these preliminary descriptive remarks, we now turn to the rather abstract but very precise ideas they are intended to motivate.

A *function* consists of three objects: two non-empty sets X and Y (which may be equal, but need not be) and a rule f which assigns to each element x in X a single fully determined element y in Y. The y which corresponds in this way to a given x is usually written $f(x)$, and is called the *image* of x under the rule f, or the *value* of f at the element x. This

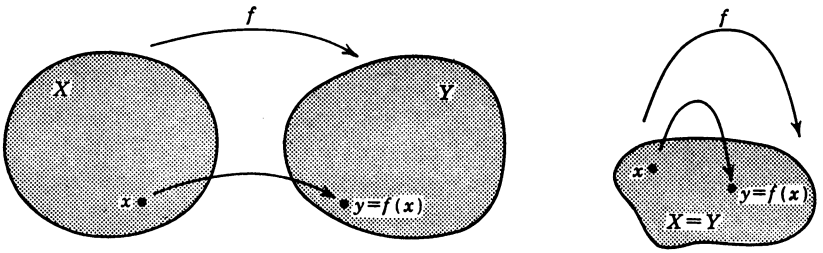

Fig. 6. A way of visualizing mappings.

notation is supposed to be suggestive of the idea that the rule f takes the element x and does something to it to produce the element $y = f(x)$. The rule f is often called a *mapping*, or *transformation*, or *operator*, to amplify this concept of it. We then think of f as mapping x's to y's, or transforming x's into y's, or operating on x's to produce y's. The set X is called the *domain* of the function, and the set of all $f(x)$'s for all x's in X is called its *range*. A function whose range consists of just one element is called a *constant function*.

We often denote by $f: X \to Y$ the function with rule f, domain X, and range contained in Y. This notation is useful because the essential parts of the function are displayed in a manner which emphasizes that it is a composite object, the central thing being the rule or mapping f. Figure 6 gives a convenient way of picturing this function. On the left, X and Y are different sets, and on the right, they are equal—in which case we usually refer to f as a mapping of X into itself. If it is clear from the context what the sets X and Y are, or if there is no real need to specify them explicitly, it is common practice to identify the function $f: X \to Y$ with the rule f, and to speak of f alone as if it were the function under consideration (without mentioning the sets X and Y).

It sometimes happens that two perfectly definite sets X and Y are under discussion and that a mapping of X into Y arises which has no natural symbol attached to it. If there is no necessity to invent a

symbol for this mapping, and if it is quite clear what the mapping is, it is often convenient to designate it by $x \to y$. Accordingly, the function $y = x^2$ mentioned at the beginning of this section can be written as $x \to x^2$ or $x \to y$ (where y is understood to be the square of x).

A function f is called an *extension* of a function g (and g is called a *restriction* of f) if the domain of f contains the domain of g and $f(x) = g(x)$ for each x in the domain of g.

Most of mathematical analysis, both classical and modern, deals with functions whose values are real numbers or complex numbers.

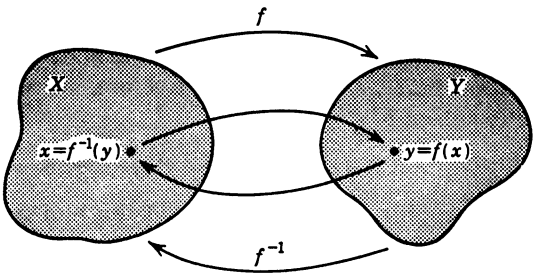

Fig. 7. The inverse of a mapping.

This is also true of those parts of topology which are concerned with the foundations of analysis. If the range of a function consists of real numbers, we call it a *real function;* similarly, a *complex function* is one whose range consists of complex numbers. Obviously, every real function is also complex. We lay very heavy emphasis on real and complex functions throughout our work.

As a matter of usage, we generally prefer to reserve the term *function* for real or complex functions and to speak of *mappings* when dealing with functions whose values are not necessarily numbers.

Consider a mapping $f: X \to Y$. When we call f a mapping of X *into* Y, we mean to suggest by this that the elements $f(x)$—as x varies over all the elements of X—need not fill up Y; but if it definitely does happen that the range of f equals Y, or if we specifically want to assume this, then we call f a mapping of X *onto* Y. If two different elements in X always have different images under f, then we call f a *one-to-one* mapping of X into Y. If $f: X \to Y$ is both onto and one-to-one, then we can define its *inverse mapping* $f^{-1}: Y \to X$ as follows: for each y in Y, we find that unique element x in X such that $f(x) = y$ (x exists and is unique since f is onto and one-to-one); we then define x to be $f^{-1}(y)$. The equation $x = f^{-1}(y)$ is the result of solving $y = f(x)$ for x in just the same way as $x = \log y$ is the result of solving $y = e^x$ for x. Figure 7 illustrates the concept of the inverse of a mapping.

If f is a one-to-one mapping of X onto Y, it will sometimes be convenient to subordinate the conception of f as a mapping sending x's over to y's and to emphasize its role as a link between x's and y's. Each x has linked to it (or has corresponding to it) precisely one $y = f(x)$; and, turning the situation around, each y has linked to it (or has corresponding to it) exactly one $x = f^{-1}(y)$. When we focus our attention on this aspect of a mapping which is one-to-one onto, we usually call it a *one-to-one correspondence.* Thus f is a one-to-one correspondence between X and Y, and f^{-1} is a one-to-one correspondence between Y and X.

Now consider an arbitrary mapping $f: X \to Y$. The mapping f, which sends each element of X over to an element of Y, induces the following two important *set mappings.* If A is a subset of X, then its *image* $f(A)$ is the subset of Y defined by

$$f(A) = \{f(x) : x \in A\},$$

and our first set mapping is that which sends each A over to its corresponding $f(A)$. Similarly, if B is a subset of Y, then its *inverse image* $f^{-1}(B)$ is the subset of X defined by

$$f^{-1}(B) = \{x : f(x) \in B\},$$

and the second set mapping pulls each B back to its corresponding $f^{-1}(B)$. It is often essential for us to know how these set mappings behave with respect to set inclusion and operations on sets. We develop most of their significant features in the following two paragraphs.

The main properties of the first set mapping are:

$$\begin{gathered} f(\emptyset) = \emptyset; \qquad f(X) \subseteq Y; \\ A_1 \subseteq A_2 \Rightarrow f(A_1) \subseteq f(A_2); \\ f(\cup_i A_i) = \cup_i f(A_i); \\ f(\cap_i A_i) \subseteq \cap_i f(A_i). \end{gathered} \tag{1}$$

The reader should convince himself of the truth of these statements. For instance, to prove (1) we would have to prove first that $f(\cup_i A_i)$ is a subset of $\cup_i f(A_i)$, and second that $\cup_i f(A_i)$ is a subset of $f(\cup_i A_i)$. A proof of the first of these set inclusions might run as follows: an element in $f(\cup_i A_i)$ is the image of some element in $\cup_i A_i$, therefore it is the image of an element in some A_i, therefore it is in some $f(A_i)$, and so finally it is in $\cup_i f(A_i)$. The irregularities and gaps which the reader will notice in the above statements are essential features of this set mapping. For example, the image of an intersection need not equal the intersection of the images, because two disjoint sets can easily have images which are not disjoint. Furthermore, without special assumptions (see Problem 6) nothing can be said about the relation between $f(A)'$ and $f(A')$.

The second set mapping is much better behaved. Its properties are satisfyingly complete, and can be stated as follows:

$$f^{-1}(\emptyset) = \emptyset; \qquad f^{-1}(Y) = X;$$
$$B_1 \subseteq B_2 \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2);$$
$$f^{-1}(\cup_i B_i) = \cup_i f^{-1}(B_i); \tag{2}$$
$$f^{-1}(\cap_i B_i) = \cap_i f^{-1}(B_i); \tag{3}$$
$$f^{-1}(B') = f^{-1}(B)'. \tag{4}$$

Again, the reader should verify each of these statements for himself.

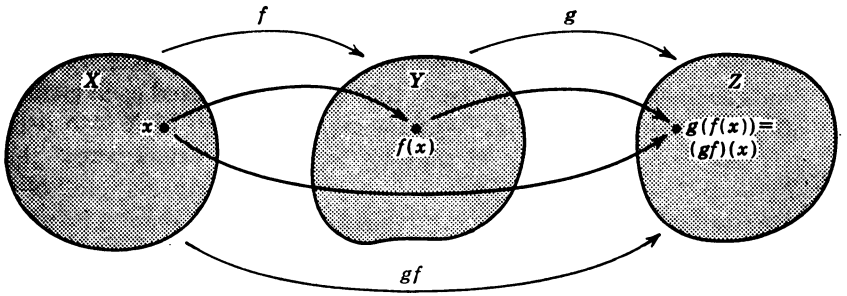

Fig. 8. Multiplication of mappings.

We discuss one more concept in this section, that of the *multiplication* (or *composition*) of mappings. If $y = f(x) = x^2 + 1$ and

$$z = g(y) = \sin y,$$

then these two functions can be put together to form a single function defined by $z = (gf)(x) = g(f(x)) = g(x^2 + 1) = \sin(x^2 + 1)$. One of the most important tools of calculus (the chain rule) explains how to differentiate functions of this kind. This manner of multiplying functions together is of basic importance for us as well, and we formulate it in general as follows. Suppose that $f: X \to Y$ and $g: Y \to Z$ are any two mappings. We define the *product* of these mappings, denoted by $gf: X \to Z$, by $(gf)(x) = g(f(x))$. In words: an element x in X is taken by f to the element $f(x)$ in Y, and then g maps $f(x)$ to $g(f(x))$ in Z. Figure 8 is a picture of this process. We observe that the two mappings involved here are not entirely arbitrary, for the set Y which contains the range of the first equals the domain of the second. More generally, the product of two mappings is meaningful whenever the range of the first is contained in the domain of the second. We have regarded f as the first mapping and g as the second, and in forming their product gf, their symbols have gotten turned around. This is a rather unpleasant phenomenon, for which we blame the occasional perversity of mathematical symbols. Perhaps it will help the reader to keep this straight

in his mind if he will remember to read the product gf from right to left: first apply f, then g.

Problems

1. Two mappings $f: X \to Y$ and $g: X \to Y$ are said to be *equal* (and we write this $f = g$) if $f(x) = g(x)$ for every x in X. Let f, g, and h be any three mappings of a non-empty set X into itself, and show that multiplication of mappings is associative in the sense that $f(gh) = (fg)h$.
2. Let X be a non-empty set. The *identity mapping* i_X on X is the mapping of X onto itself defined by $i_X(x) = x$ for every x. Thus i_X sends each element of X to itself; that is, it leaves fixed each element of X. Show that $fi_X = i_X f = f$ for any mapping f of X into itself. If f is one-to-one onto, so that its inverse f^{-1} exists, show that $ff^{-1} = f^{-1}f = i_X$. Show further that f^{-1} is the only mapping of X into itself which has this property; that is, show that if g is a mapping of X into itself such that $fg = gf = i_X$, then $g = f^{-1}$ (*hint:* $g = gi_X = g(ff^{-1}) = (gf)f^{-1} = i_X f^{-1} = f^{-1}$, or
$$g = i_X g = (f^{-1}f)g = f^{-1}(fg) = f^{-1}i_X = f^{-1}).$$
3. Let X and Y be non-empty sets and f a mapping of X into Y. Show the following:
   (*a*) f is one-to-one $\Leftrightarrow$ there exists a mapping g of Y into X such that $gf = i_X$;
   (*b*) f is onto $\Leftrightarrow$ there exists a mapping h of Y into X such that $fh = i_Y$.
4. Let X be a non-empty set and f a mapping of X into itself. Show that f is one-to-one onto $\Leftrightarrow$ there exists a mapping g of X into itself such that $fg = gf = i_X$. If there exists a mapping g with this property, then there is only one such mapping. Why?
5. Let X be a non-empty set, and let f and g be one-to-one mappings of X onto itself. Show that fg is also a one-to-one mapping of X onto itself and that $(fg)^{-1} = g^{-1}f^{-1}$.
6. Let X and Y be non-empty sets and f a mapping of X into Y. If A and B are, respectively, subsets of X and Y, show the following:
   (*a*) $ff^{-1}(B) \subseteq B$, and $ff^{-1}(B) = B$ is true for all $B \Leftrightarrow f$ is onto;
   (*b*) $A \subseteq f^{-1}f(A)$, and $A = f^{-1}f(A)$ is true for all $A \Leftrightarrow f$ is one-to-one;
   (*c*) $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$ is true for all A_1 and $A_2 \Leftrightarrow f$ is one-to-one;
   (*d*) $f(A)' \subseteq f(A')$ is true for all $A \Leftrightarrow f$ is onto;
   (*e*) if f is onto—so that $f(A)' \subseteq f(A')$ is true for all A—then $f(A)' = f(A')$ is true for all $A \Leftrightarrow f$ is also one-to-one.

4. PRODUCTS OF SETS

We shall often have occasion to weld together the sets of a given class into a single new set called their *product* (or their *Cartesian product*). The ancestor of this concept is the coordinate plane of analytic geometry, that is, a plane equipped with the usual rectangular coordinate system. We give a brief description of this fundamental idea with a view to paving the way for our discussion of products of sets in general.

First, a few preliminary comments about the *real line.* We have already used this term several times without any explanation, and of course what we mean by it is an ordinary geometric straight line (see Fig. 9) whose points have been identified with—or coordinatized by—the

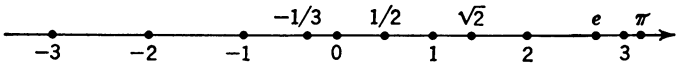

Fig. 9. The real line.

set R of all real numbers. We use the letter R to denote the real line as well as the set of all real numbers, and we often speak of real numbers as if they were points on the real line, and of points on the real line as if they were real numbers. Let no one be deceived into thinking that the real line is a simple thing, for its structure is exceedingly intricate. Our present view of it, however, is as naive and uncomplicated as the picture of it given in Fig. 9. Generally speaking, we assume that the reader is familiar with the simpler properties of the real line—those relating to inequalities (see Problem 1-2) and the basic algebraic operations of addition, subtraction, multiplication, and division. One of the most significant facts about the real number system is perhaps less well known. This is the so-called *least upper bound property,* which asserts that every non-empty set of real numbers which has an upper bound has a least upper bound. It is an easy consequence of this that every non-empty set of real numbers which has a lower bound has a greatest lower bound. All these matters can be developed rigorously on the basis of a small number of axioms, and detailed treatments can often be found in books on elementary abstract algebra.

To construct the coordinate plane, we now proceed as follows. We take two identical replicas of the real line, which we call the *x axis* and the *y axis,* and paste them on a plane at right angles to one another in such a way that they cross at the zero point on each. The usual picture is given in Fig. 10. Now let P be a point in the plane. We project P perpendicularly onto points P_x and P_y on the axes. If x and y are the coordinates of P_x and P_y on their respective axes, this process

leads us from the point P to the uniquely determined ordered pair (x,y) of real numbers, where x and y are called the *x coordinate* and *y coordinate* of P. We can reverse the process, and, starting with the ordered pair of real numbers, we can recapture the point. This is the manner in which we establish the familiar one-to-one correspondence between points P in the plane and ordered pairs (x,y) of real numbers. In fact, we think of a point in the plane (which is a geometric object) and its corresponding ordered pair of real numbers (which is an algebraic object) as being—to all intents and purposes—*identical with one another.* The essence of analytic geometry lies in the possibility of exploiting this identification by using algebraic tools in geometric arguments and giving geometric interpretations to algebraic calculations.

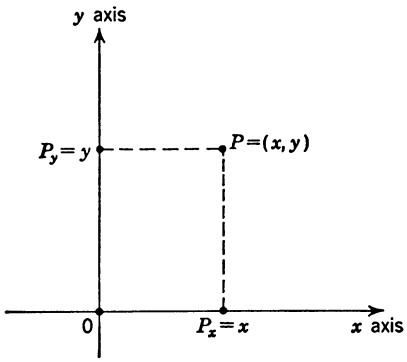

Fig. 10. The coordinate plane.

The conventional attitude toward the coordinate plane in analytic geometry is that the geometry is the focus of interest and the algebra of ordered pairs is only a convenient tool. Here we reverse this point of view. For us, the *coordinate plane* is defined to be the set of all ordered pairs (x,y) of real numbers. We can satisfy our desire for visual images by using Fig. 10 as a picture of this set and by calling such an ordered pair a point, but this geometric language is more a convenience than a necessity.

Our notation for the coordinate plane is $R \times R$, or R^2. This symbolism reflects the idea that the coordinate plane is the result of "multiplying together" two replicas of the real line R.

It is perhaps necessary to comment on one possible source of misunderstanding. When we speak of R^2 as a plane, we do so only to establish an intuitive bond with the reader's previous experience in analytic geometry. Our present attitude is that R^2 is a pure set and has no structure whatever, because no structure has yet been assigned to it. We remarked earlier (with deliberate vagueness) that a space is a set to which has been added some kind of algebraic or geometric structure. In Sec. 15 we shall convert the *set* R^2 into the *space* of analytic geometry by defining the distance between any two points (x_1,y_1) and (x_2,y_2) to be

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

This notion of distance endows the set R^2 with a certain "spatial" character, which we shall recognize by calling the resulting space the *Euclidean plane* instead of the coordinate plane.

We assume that the reader is fully acquainted with the way in which the set C of all complex numbers can be identified (as a set) with the coordinate plane R^2. If z is a complex number, and if z has the standard form $x + iy$ where x and y are real numbers, then we identify z with the ordered pair (x,y), and thus with an element of R^2. The complex numbers, however, are much more than merely a set. They constitute a number system, with operations of addition, multiplication, conjugation, etc. When the coordinate plane R^2 is thought of as consisting of complex numbers and is enriched by the algebraic structure it acquires in this way, it is called the *complex plane.* The letter C is used to denote either the set of all complex numbers or the complex plane. We shall make a space out of the complex plane in Sec. 9.

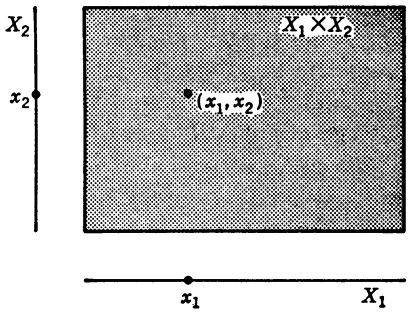

Fig. 11. A way of visualizing $X_1 \times X_2$.

Suppose now that X_1 and X_2 are any two non-empty sets. By analogy with our above discussion, their *product* $X_1 \times X_2$ is defined to be the set of all ordered pairs (x_1,x_2), where x_1 is in X_1 and x_2 is in X_2. In spite of the arbitrary nature of X_1 and X_2, their product can be represented by a picture (see Fig. 11) which is loosely similar to the usual picture of the coordinate plane. The term *product* is applied to this set, and it is thought of as the result of "multiplying together" X_1 and X_2, for the following reason: if X_1 and X_2 are finite sets with m and n elements, then (clearly) $X_1 \times X_2$ has mn elements. If $f: X_1 \to X_2$ is a mapping with domain X_1 and range in X_2, its *graph* is that subset of $X_1 \times X_2$ which consists of all ordered pairs of the form $(x_1,f(x_1))$. We observe that this is an appropriate generalization of the concept of the graph of a function as it occurs in elementary mathematics.

This definition of the product of two sets extends easily to the case of n sets for any positive integer n. If $X_1, X_2, \ldots, X_n$ are non-empty sets, then their *product* $X_1 \times X_2 \times \cdots \times X_n$ is the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$, where x_i is in X_i for each subscript i. If the X_i's are all replicas of a single set X, that is, if

$$X_1 = X_2 = \cdots = X_n = X,$$

then their product is usually denoted by the symbol X^n.

These ideas specialize directly to yield the important sets R^n and C^n. R^1 is just R, the real line, and R^2 is the coordinate plane. R^3—the set of all ordered triples of real numbers—is the set which underlies solid analytic geometry, and we assume that the reader is familiar with

the manner in which this set arises, through the introduction of a rectangular coordinate system into ordinary three-dimensional space. We can draw pictures here just as in the case of the coordinate plane, and we can use geometric language as much as we please, but it must be understood that the mathematics of this set is the mathematics of ordered triples of real numbers and that the pictures are merely an aid to the intuition. Once we fully grasp this point of view, there is no difficulty whatever in advancing at once to the study of the set R^n of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers for any positive integer n. It is quite true that when n is greater than 3 it is no longer possible to draw the same kinds of intuitively rich pictures, but at worst this is merely an inconvenience. We can (and do) continue to use suggestive geometric language, so all is not lost. The set C^n is defined similarly: it is the set of all ordered n-tuples $(z_1, z_2, \ldots, z_n)$ of complex numbers. Each of the sets R^n and C^n plays a prominent part in our later work.

We emphasized above that for the present the coordinate plane is to be considered as merely a set, and not a space. Similar remarks apply to R^n and C^n. In due course (in Sec. 15) we shall impart form and content to each of these sets by suitable definitions. We shall convert them into the *Euclidean* and *unitary n-spaces* which underlie and motivate so many developments in modern pure mathematics, and we shall explore some aspects of their algebraic and topological structure to the very last pages of this book. But as of now—and this is the point we insist on—neither one of these sets has any structure at all.

As the reader doubtless suspects, it is not enough that we consider only products of finite classes of sets. The needs of topology compel us to extend these ideas to arbitrary classes of sets.

We defined the product $X_1 \times X_2 \times \cdots \times X_n$ to be the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$ such that x_i is in X_i for each subscript i. To see how to extend this definition, we reformulate it as follows. We have an index set I, consisting of the integers from 1 to n, and corresponding to each index (or subscript) i we have a non-empty set X_i. The n-tuple $(x_1, x_2, \ldots, x_n)$ is simply a function (call it x) defined on the index set I, with the restriction that its value $x(i) = x_i$ is an element of the set X_i for each i in I. Our point of view here is that the function x is completely determined by, and is essentially equivalent to, the array $(x_1, x_2, \ldots, x_n)$ of its values.

The way is now open for the definition of products in their full generality. Let $\{X_i\}$ be a non-empty class of non-empty sets, indexed by the elements i of an index set I. The sets X_i need not be different from one another; indeed, it may happen that they are all identical replicas of a single set, distinguished only by different indices. The *product* of the sets X_i, written $P_{i \in I} X_i$, is defined to be the set of all functions x defined on I such that $x(i)$ is an element of the set X_i for

each index i. We call X_i the *ith coordinate set.* When there can be no misunderstanding about the index set, the symbol $P_{i\in I}\, X_i$ is often abbreviated to P_iX_i. The definition we have just given requires that each coordinate set be non-empty before the product can be formed. It will be useful if we extend this definition slightly by agreeing that if any of the X_i's are empty, then P_iX_i is also empty.

This approach to the idea of the product of a class of sets, by means of functions defined on the index set, is useful mainly in giving the definition. In practice, it is much more convenient to use the subscript notation x_i instead of the function notation $x(i)$. We then interpret the product P_iX_i as made up of elements x, each of which is specified by the exhibited array $\{x_i\}$ of its values in the respective coordinate sets X_i. We call x_i the *ith coordinate* of the element $x = \{x_i\}$.

The mapping p_i of the product P_iX_i onto its ith coordinate set X_i which is defined by $p_i(x) = x_i$—that is, the mapping whose value at an arbitrary element of the product is the ith coordinate of that element —is called the *projection* onto the ith coordinate set. The projection p_i selects the ith coordinate of each element in its domain. There is clearly one projection for each element of the index set I, and the set of all projections plays an important role in the general theory of topological spaces.

Problems

1. The graph of a mapping $f: X \to Y$ is a subset of the product $X \times Y$. What properties characterize the graphs of mappings among all subsets of $X \times Y$?
2. Let X and Y be non-empty sets. If A_1 and A_2 are subsets of X, and B_1 and B_2 subsets of Y, show the following:

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2);$$
$$(A_1 \times B_1) - (A_2 \times B_2) = (A_1 - A_2) \times (B_1 - B_2) \cup (A_1 \cap A_2) \times (B_1 - B_2) \cup (A_1 - A_2) \times (B_1 \cap B_2).$$

3. Let X and Y be non-empty sets, and let $\mathbf{A}$ and $\mathbf{B}$ be rings of subsets of X and Y, respectively. Show that the class of all finite unions of sets of the form $A \times B$ with $A \in \mathbf{A}$ and $B \in \mathbf{B}$ is a ring of subsets of $X \times Y$.

5. PARTITIONS AND EQUIVALENCE RELATIONS

In the first part of this section we consider a non-empty set X, and we study decompositions of X into non-empty subsets which fill it out

and have no elements in common with one another. We give special attention to the tools (equivalence relations) which are normally used to generate such decompositions.

A *partition* of X is a disjoint class $\{X_i\}$ of non-empty subsets of X whose union is the full set X itself. The X_i's are called the *partition sets.* Expressed somewhat differently, a partition of X is the result of splitting it, or subdividing it, into non-empty subsets in such a way that each element of X belongs to one and only one of the given subsets.

If X is the set $\{1, 2, 3, 4, 5\}$, then $\{1, 3, 5\}$, $\{2, 4\}$ and $\{1, 2, 3\}$, $\{4, 5\}$ are two different partitions of X. If X is the set R of all real numbers, then we can partition X into the set of all rationals and the set of all irrationals, or into the infinitely many closed-open intervals of the form $[n, n + 1)$ where n is an integer. If X is the set of all points in the coordinate plane, then we can partition X in such a way that each partition set consists of all points with the same x coordinate (vertical lines), or so that each partition set consists of all points with the same y coordinate (horizontal lines).

Other partitions of each of these sets will readily occur to the reader. In general, there are many different ways in which any given set can be partitioned. These manufactured examples are admittedly rather uninspiring and serve only to make our ideas more concrete. Later in this section we consider some others which are more germane to our present purposes.

A *binary relation* in the set X is a mathematical symbol or verbal phrase, which we denote by R in this paragraph, such that for each ordered pair (x,y) of elements of X the statement $x \mathrm{R} y$ is meaningful, in the sense that it can be classified definitely as true or false. For such a binary relation, $x \mathrm{R} y$ symbolizes the assertion that x *is* related by R to y, and $x \not\mathrm{R} y$ the negation of this, namely, the assertion that x *is not* related by R to y. Many examples of binary relations can be given, some familiar and others less so, some mathematical and others not. For instance, if X is the set of all integers and R is interpreted to mean "is less than," which of course is usually denoted by the symbol $<$, then we clearly have $4 < 7$ and $5 \not< 2$. We have been speaking of binary relations, which are so named because they apply only to ordered pairs of elements, rather than to ordered triples, etc. In our work we drop the qualifying adjective and speak simply of a *relation* in X, since we shall have occasion to consider only relations of this kind.[1]

We now assume that a partition of our non-empty set X is given,

[1] Some writers prefer to regard a relation R in X as a subset R of $X \times X$. From this point of view, $x \mathrm{R} y$ and $x \not\mathrm{R} y$ are simply equivalent ways of writing $(x,y) \in \mathrm{R}$ and $(x,y) \notin \mathrm{R}$. This definition has the advantage of being more tangible than ours, and the disadvantage that few people really think of a relation in this way.

and we associate with this partition a relation in X. This relation is defined in the following way: we say that x is *equivalent* to y and write this $x \sim y$ (the symbol $\sim$ is pronounced "wiggle"), if x and y belong to the same partition set. It is obvious that the relation $\sim$ has the following properties:

(1) $x \sim x$ for every x (*reflexivity*);

(2) $x \sim y \Rightarrow y \sim x$ (*symmetry*);

(3) $x \sim y$ and $y \sim z \Rightarrow x \sim z$ (*transitivity*).

This particular relation in X arose in a special way, in connection with a given partition of X, and its properties are immediate consequences of its definition. Any relation whatever in X which possesses these three properties is called an *equivalence relation* in X.

We have just seen that each partition of X has associated with it a natural equivalence relation in X. We now reverse the situation and show that a given equivalence relation in X determines a natural partition of X.

Let $\sim$ be an equivalence relation in X; that is, assume that it is reflexive, symmetric, and transitive in the sense described above. If x is an element of X, the subset of X defined by $[x] = \{y : y \sim x\}$ is called the *equivalence set* of x. The equivalence set of x is thus the set of all elements which are equivalent to x. We show that the class of all distinct equivalence sets forms a partition of X. By reflexivity, $x \in [x]$ for each element x in X, so each equivalence set is non-empty and their union is X. It remains to be shown that any two equivalence sets $[x_1]$ and $[x_2]$ are either disjoint or identical. We prove this by showing that if $[x_1]$ and $[x_2]$ are not disjoint, then they must be identical. Suppose that $[x_1]$ and $[x_2]$ are not disjoint; that is, suppose that they have a common element z. Since z belongs to both equivalence sets, $z \sim x_1$ and $z \sim x_2$, and by symmetry, $x_1 \sim z$. Let y be any element of $[x_1]$, so that $y \sim x_1$. Since $y \sim x_1$ and $x_1 \sim z$, transitivity shows that $y \sim z$. By another application of transitivity, $y \sim z$ and $z \sim x_2$ imply that $y \sim x_2$, so that y is in $[x_2]$. Since y was chosen arbitrarily in $[x_1]$, we see by this that $[x_1] \subseteq [x_2]$. The same reasoning shows that $[x_2] \subseteq [x_1]$, and from this we conclude (see the last paragraph of Sec. 1) that $[x_1] = [x_2]$.

The above discussion demonstrates that there is no real distinction (other than a difference in language) between partitions of a set and equivalence relations in the set. If we start with a partition, we get an equivalence relation by regarding elements as equivalent if they belong to the same partition set, and if we start with an equivalence relation, we get a partition by grouping together into subsets all elements which are equivalent to one another. We have here a single mathematical idea, which we have been considering from two different points of view, and the approach we choose in any particular application depends entirely

on our own convenience. In practice, it is almost invariably the case that we use equivalence relations (which are usually easy to define) to obtain partitions (which are sometimes difficult to describe fully).

We now turn to several of the more important simple examples of equivalence relations.

Let I be the set of all integers. If a and b are elements of this set, we write $a = b$ (and say that a *equals* b) if a and b are the same integer. Thus $2 + 3 = 5$ means that the expressions on the left and right are simply different ways of writing the same integer. It is apparent that $=$ used in this sense is an equivalence relation in the set I:

(1) $a = a$ for every a;

(2) $a = b \Rightarrow b = a$;

(3) $a = b$ and $b = c \Rightarrow a = c$.

Clearly, each equivalence set consists of precisely one integer.

Another familiar example is the relation of equality commonly used for fractions. We remind the reader that, strictly speaking, a fraction is merely a symbol of the form a/b, where a and b are integers and b is not zero. The fractions $\frac{2}{3}$ and $\frac{4}{6}$ are obviously not identical, but nevertheless we consider them to be equal. In general, we say that two fractions a/b and c/d are *equal*, written $a/b = c/d$, if ad and bc are equal as integers in the usual sense (see the above paragraph). We leave it to the reader to show that this is an equivalence relation in the set of all fractions. An equivalence set of fractions is what we call a *rational number*. Everyday usage ignores the distinction between fractions and rational numbers, but it is important to recognize that from the strict point of view it is the rational numbers (and not the fractions) which form part of the real number system.

Our final example has a deeper significance, for it provides us with the basic tool for our work of the next two sections.

For the remainder of this section we consider a relation between pairs of non-empty sets, and each set mentioned (whether we say so explicitly or not) is assumed to be non-empty. If X and Y are two sets, we say that X is *numerically equivalent* to Y if there exists a one-to-one correspondence between X and Y, i.e., if there exists a one-to-one mapping of X onto Y. This relation is reflexive, since the identity mapping $i_X : X \to X$ is one-to-one onto; it is symmetric, since if $f : X \to Y$ is one-to-one onto, then its inverse mapping $f^{-1} : Y \to X$ is also one-to-one onto; and it is transitive, since if $f : X \to Y$ and $g : Y \to Z$ are one-to-one onto, then $gf : X \to Z$ is also one-to-one onto. Numerical equivalence has all the properties of an equivalence relation, and if we consider it as an equivalence relation in the class of all non-empty subsets of some universal set U, it groups together into equivalence sets all those subsets of U which have the *same number of elements*. After we state and prove the

following very useful but rather technical theorem, we shall continue in Secs. 6 and 7 with an exploration of the implications of these ideas.

The theorem we have in mind—the *Schroeder-Bernstein theorem*—is the following: *if X and Y are two sets each of which is numerically equivalent to a subset of the other, then all of X is numerically equivalent to all of Y.* There are several proofs of this classic theorem, some of which are quite difficult. The very elegant proof we give is essentially due to Birkhoff and MacLane.

Now for the proof. We assume that $f: X \to Y$ is a one-to-one mapping of X into Y, and that $g: Y \to X$ is a one-to-one mapping of Y into X. Our task is to produce a mapping $F: X \to Y$ which is one-to-one *onto*. We may assume that neither f nor g is onto, since if f is, we can define F to be f, and if g is, we can define F to be g^{-1}. Since both f and g are one-to-one, it is permissible to use the mappings f^{-1} and g^{-1} as long as we clearly understand that f^{-1} is defined only on $f(X)$ and g^{-1} only on $g(Y)$. We obtain the mapping F by splitting both X and Y into subsets which we characterize in terms of the ancestry of their elements. Let x be an element of X. We apply g^{-1} to it (if we can) to get the element $g^{-1}(x)$ in Y. If $g^{-1}(x)$ exists, we call it the first ancestor of x. The element x itself we call the zeroth ancestor of x. We now apply f^{-1} to $g^{-1}(x)$ if we can, and if $(f^{-1}g^{-1})(x)$ exists, we call it the second ancestor of x. We now apply g^{-1} to $(f^{-1}g^{-1})(x)$ if we can, and if $(g^{-1}f^{-1}g^{-1})(x)$ exists, we call it the third ancestor of x. As we continue this process of tracing back the ancestry of x, it becomes apparent that there are three possibilities. (1) x has infinitely many ancestors. We denote by X_i the subset of X which consists of all elements with infinitely many ancestors. (2) x has an even number of ancestors; this means that x has a last ancestor (that is, one which itself has no first ancestor) in X. We denote by X_e the subset of X consisting of all elements with an even number of ancestors. (3) x has an odd number of ancestors; this means that x has a last ancestor in Y. We denote by X_o the subset of X which consists of all elements with an odd number of ancestors. The three sets X_i, X_e, X_o form a disjoint class whose union is X. We decompose Y in just the same way into three subsets Y_i, Y_e, Y_o. It is easy to see that f maps X_i onto Y_i and X_e onto Y_o, and that g^{-1} maps X_o onto Y_e; and we complete the proof by defining F in the following piecemeal manner:

$$F(x) = \begin{cases} f(x) & \text{if } x \in X_i \cup X_e, \\ g^{-1}(x) & \text{if } x \in X_o. \end{cases}$$

We attempt to illustrate these ideas in Fig. 12. Here we present two replicas of the situation: on the left, X and Y are represented by the vertical lines, and f and g by the lines slanting down to the right and

left; and on the right, we schematically trace the ancestry of three elements in X, of which x_1 has no first ancestor, x_2 has a first and second ancestor, and x_3 has a first, second, and third ancestor.

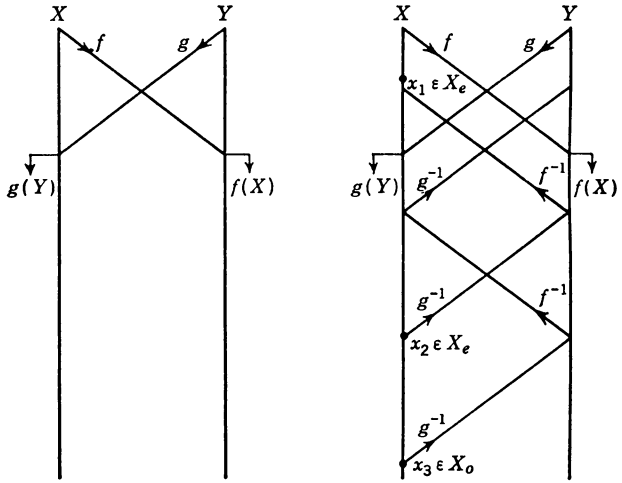

Fig. 12. The proof of the Schroeder-Bernstein theorem.

The Schroeder-Bernstein theorem has great theoretical and practical significance. Its main value for us lies in its role as a tool by means of which we can prove numerical equivalence with a minimum of effort for many specific sets. We put it to work in Sec. 7.

Problems

1. Let f: $X \to Y$ be an arbitrary mapping. Define a relation in X as follows: $x_1 \sim x_2$ means that $f(x_1) = f(x_2)$. Show that this is an equivalence relation and describe the equivalence sets.
2. In the set R of all real numbers, let $x \sim y$ mean that $x - y$ is an integer. Show that this is an equivalence relation and describe the equivalence sets.
3. Let I be the set of all integers, and let m be a fixed positive integer. Two integers a and b are said to be *congruent* modulo m—symbolized by $a \equiv b \pmod{m}$—if $a - b$ is exactly divisible by m, i.e., if $a - b$ is an integral multiple of m. Show that this is an equivalence relation, describe the equivalence sets, and state the number of distinct equivalence sets.
4. Decide which ones of the three properties of reflexivity, symmetry, and transitivity are true for each of the following relations in the set

of all positive integers: $m \leq n$, $m < n$, m divides n. Are any of these equivalence relations?

5. Give an example of a relation which is (*a*) reflexive but not symmetric or transitive; (*b*) symmetric but not reflexive or transitive; (*c*) transitive but not reflexive or symmetric; (*d*) reflexive and symmetric but not transitive; (*e*) reflexive and transitive but not symmetric; (*f*) symmetric and transitive but not reflexive.
6. Let X be a non-empty set and $\sim$ a relation in X. The following purports to be a proof of the statement that if this relation is symmetric and transitive, then it is necessarily reflexive: $x \sim y \Rightarrow y \sim x$; $x \sim y$ and $y \sim x \Rightarrow x \sim x$; therefore $x \sim x$ for every x. In view of Problem 5*f*, this cannot be a valid proof. What is the flaw in the reasoning?
7. Let X be a non-empty set. A relation $\sim$ in X is called *circular* if $x \sim y$ and $y \sim z \Rightarrow z \sim x$, and *triangular* if $x \sim y$ and $x \sim z \Rightarrow y \sim z$. Prove that a relation in X is an equivalence relation $\Leftrightarrow$ it is reflexive and circular $\Leftrightarrow$ it is reflexive and triangular.

6. COUNTABLE SETS

The subject of this section and the next—*infinite cardinal numbers*—lies at the very foundation of modern mathematics. It is a vital instrument in the day-to-day work of many mathematicians, and we shall make extensive use of it ourselves. This theory, which was created by the German mathematician Cantor, also has great aesthetic appeal, for it begins with ideas of extreme simplicity and develops through natural stages into an elaborate and beautiful structure of thought. In the course of our discussion we shall answer questions which no one before Cantor's time thought to ask, and we shall ask a question which no one can answer to this day.

Without further ado, we can say that *cardinal numbers* are those used in counting, such as the positive integers (or natural numbers) 1, 2, 3, . . . familiar to us all. But there is much more to the story than this.

The act of counting is undoubtedly one of the oldest of human activities. Men probably learned to count in a crude way at about the same time as they began to develop articulate speech. The earliest men who lived in communities and domesticated animals must have found it necessary to record the number of goats in the village herd by means of a pile of stones or some similar device. If the herd was counted in each night by removing one stone from the pile for each goat accounted for, then stones left over would have indicated strays, and herdsmen would have gone out to search for them. Names for numbers and symbols for

them, like our 1, 2, 3, . . . , would have been superfluous. The simple and yet profound idea of a one-to-one correspondence between the stones and the goats would have fully met the needs of the situation.

In a manner of speaking, we ourselves use the infinite set

$$N = \{1, 2, 3, \ldots\}$$

of all positive integers as a "pile of stones." We carry this set around with us as part of our intellectual equipment. Whenever we want to count a set, say, a stack of dollar bills, we start through the set N and tally off one bill against each positive integer as we come to it. The last number we reach, corresponding to the last bill, is what we call the number of bills in the stack. If this last number happens to be 10, then "10" is our symbol for the number of bills in the stack, as it also is for the number of our fingers, and for the number of our toes, and for the number of elements in any set which can be put into one-to-one correspondence with the finite set $\{1, 2, \ldots, 10\}$. Our procedure is slightly more sophisticated than that of the primitive savage. We have the symbols 1, 2, 3, . . . for the numbers which arise in counting; we can record them for future use, and communicate them to other people, and manipulate them by the operations of arithmetic. But the underlying idea, that of the one-to-one correspondence, remains the same for us as it probably was for him.

The positive integers are adequate for the purpose of counting any non-empty finite set, and since outside of mathematics all sets appear to be of this kind, they suffice for all non-mathematical counting. But in the world of mathematics we are obliged to consider many infinite sets, such as the set of all positive integers itself, the set of all integers, the set of all rational numbers, the set of all real numbers, the set of all points in a plane, and so on. It is often important to be able to count such sets, and it was Cantor's idea to do this, and to develop a theory of infinite cardinal numbers, by means of one-to-one correspondences.

In comparing the sizes of two sets, the basic concept is that of numerical equivalence as defined in the previous section. We recall that two non-empty sets X and Y are said to be numerically equivalent if there exists a one-to-one mapping of one onto the other, or—and this amounts to the same thing—if there can be found a one-to-one correspondence between them. To say that two non-empty finite sets are numerically equivalent is of course to say that they have the *same number of elements* in the ordinary sense. If we count one of them, we simply establish a one-to-one correspondence between its elements and a set of positive integers of the form $\{1, 2, \ldots, n\}$, and we then say that n is the *number of elements possessed by both*, or the *cardinal number of both*. The positive integers are the *finite cardinal numbers*. We encounter

many surprises as we follow Cantor and consider numerical equivalence for infinite sets.

The set $N = \{1, 2, 3, \ldots\}$ of all positive integers is obviously "larger" than the set $\{2, 4, 6, \ldots\}$ of all even positive integers, for it contains this set as a proper subset. It appears on the surface that N has "more" elements. But it is very important to avoid jumping to conclusions when dealing with infinite sets, and we must remember that our criterion in these matters is whether there exists a one-to-one correspondence between the sets (not whether one set is or is not a proper subset of the other). As a matter of fact, the pairing

$$\begin{array}{l} 1, 2, 3, \ldots, n, \ldots \\ 2, 4, 6, \ldots, 2n, \ldots \end{array}$$

serves to establish a one-to-one correspondence between these sets, in which each positive integer in the upper row is matched with the even positive integer (its double) directly below it, and these two sets must therefore be regarded as having the *same number of elements.* This is a very remarkable circumstance, for it seems to contradict our intuition and yet is based only on solid common sense. We shall see below, in Problems 6 and 7-4, that every infinite set is numerically equivalent to a proper subset of itself. Since this property is clearly not possessed by any finite set, some writers even use it as the definition of an infinite set.

In much the same way as above, we can show that N is numerically equivalent to the set of *all* even integers:

$$\begin{array}{l} 1, 2, \quad 3, 4, \quad 5, 6, \quad 7, \ldots \\ 0, 2, -2, 4, -4, 6, -6, \ldots \end{array}$$

Here our device is to start with 0 and follow each even positive integer as we come to it by its negative. Similarly, N is numerically equivalent to the set of all integers:

$$\begin{array}{l} 1, 2, \quad 3, 4, \quad 5, 6, \quad 7, \ldots \\ 0, 1, -1, 2, -2, 3, -3, \ldots \end{array}$$

It is of considerable historical interest to note that Galileo observed in the early seventeenth century that there are precisely as many perfect squares (1, 4, 9, 16, 25, etc.) among the positive integers as there are positive integers altogether. This is clear from the pairing

$$\begin{array}{l} 1, \ 2, \ 3, \ 4, \ 5, \ldots \\ 1^2, 2^2, 3^2, 4^2, 5^2, \ldots \end{array}$$

It struck him as very strange that this should be true, considering how

sparsely strewn the squares are among all the positive integers. But the time appears not to have been ripe for the exploration of this phenomenon, or perhaps he had other things on his mind; in any case, he did not follow up his idea.

These examples should make it clear that all that is really necessary in showing that an infinite set X is numerically equivalent to N is that we be able to list the elements of X, with a first, a second, a third, and so on, in such a way that it is completely exhausted by this counting off of its elements. It is for this reason that any infinite set which is numerically equivalent to N is said to be *countably infinite.* We say that a set is *countable* if it is non-empty and finite (in which case it can obviously be counted) or if it is countably infinite.

One of Cantor's earliest discoveries in his study of infinite sets was that the set of all positive rational numbers (which is very large: it contains N and a great many other numbers besides) is actually countable. We cannot list the positive rational numbers in order of size, as we can the positive integers, beginning with the smallest, then the next smallest, and so on, for there is no smallest, and between any two there are infinitely many others. We must find some other way of counting them, and following Cantor, we arrange them not in order of size, but according to the size of the sum of the numerator and denominator. We begin with all positive rationals whose numerator and denominator add up to 2: there is only one, $\frac{1}{1} = 1$. Next we list (with increasing numerators) all those for which this sum is $3: \frac{1}{2}, \frac{2}{1} = 2$. Next, all those for which this sum is $4: \frac{1}{3}, \frac{2}{2} = 1, \frac{3}{1} = 3$. Next, all those for which this sum is $5: \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1} = 4$. Next, all those for which this sum is $6: \frac{1}{5}, \frac{2}{4} = \frac{1}{2}, \frac{3}{3} = 1, \frac{4}{2} = 2, \frac{5}{1} = 5$. And so on. If we now list all these together from the beginning, omitting those already listed when we come to them, we get a sequence

$$1, \tfrac{1}{2}, 2, \tfrac{1}{3}, 3, \tfrac{1}{4}, \tfrac{2}{3}, \tfrac{3}{2}, 4, \tfrac{1}{5}, 5, \ldots$$

which contains each positive rational number once and only once. Figure 13 gives a schematic representation of this manner of listing the positive rationals. In this figure the first row contains all positive rationals with numerator 1, the second all with numerator 2, etc.; and the first column contains all with denominator 1, the second all with denominator 2, and so on. Our listing amounts to traversing this array of numbers as the arrows indicate, where of course all those numbers already encountered are left out.

It's high time that we christened the infinite cardinal number we've been discussing, and for this purpose we use the first letter of the Hebrew alphabet ($\aleph$, pronounced "aleph") with 0 as a subscript. We say that $\aleph_0$ is the number of elements in any countably infinite set. Our

complete list of cardinal numbers so far is

$$1, 2, 3, \ldots, \aleph_0.$$

We expand this list in the next section.

Suppose now that m and n are two cardinal numbers (finite or infinite). The statement that m is *less than* n (written $m < n$) is defined to mean the following: if X and Y are sets with m and n elements, then

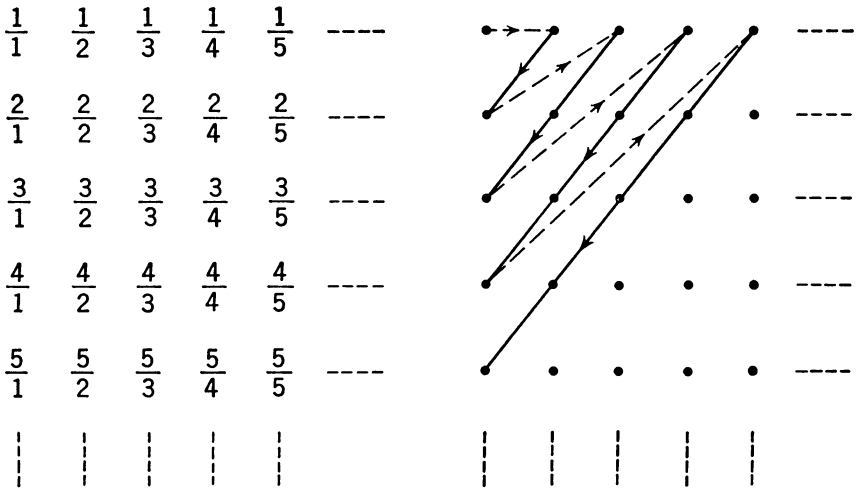

Fig. 13. A listing of the positive rationals.

(1) there exists a one-to-one mapping of X into Y, and (2) there does not exist a one-to-one mapping of X onto Y. Using this concept, it is easy to relate our cardinal numbers to one another by means of

$$1 < 2 < 3 < \cdots < \aleph_0.$$

With respect to the finite cardinal numbers, this ordering corresponds to their usual ordering as real numbers.

Problems

1. Prove that the set of all rational numbers (positive, negative, and zero) is countable. (*Hint:* see our method of showing that the set of all integers is countable.)
2. Use the idea behind Fig. 13 to prove that if $\{X_i\}$ is a countable class of countable sets, then $\cup_i X_i$ is also countable. We usually express this by saying that *any countable union of countable sets is countable.*

3. Prove that the set of all rational points in the coordinate plane R^2 (i.e., all points whose coordinates are both rational) is countable.
4. Prove that if X_1 and X_2 are countable, then $X_1 \times X_2$ is also countable.
5. Prove that if $X_1, X_2, \ldots, X_n$ are countable, where n is any positive integer, then $X_1 \times X_2 \times \cdots \times X_n$ is also countable.
6. Prove that every countably infinite set is numerically equivalent to a proper subset of itself.
7. Prove that any non-empty subset of a countable set is countable.
8. Let X and Y be non-empty sets, and f a mapping of X onto Y. If X is countable, prove that Y is also countable.

7. UNCOUNTABLE SETS

All the infinite sets we considered in the previous section were countable, so it might appear at this stage that *every* infinite set is countable. If this were true, if the end result of the analysis of infinite sets were that they are all numerically equivalent to one another, then Cantor's theory would be relatively trivial. But this is not the case, for Cantor discovered that the infinite set R of all real numbers is *not* countable—or, as we phrase it, R is *uncountable* or *uncountably infinite*. Since we customarily identify the elements of R with the points of the real line (see Sec. 4), this amounts to the assertion that the set of *all* points on the real line represents a "higher type of infinity" than that of only the integral points or only the rational points.

Cantor's proof of this is very ingenious, but it is actually quite simple. In outline the procedure is as follows: we assume that all the real numbers (in decimal form) can be listed, and in fact have been listed; then we produce a real number which cannot be in this list—thus contradicting our initial assumption that a complete listing is possible. In representing real numbers by decimals, we use the scheme of decimal expansion in which infinite chains of 9's are avoided; for instance, we write $\frac{1}{2}$ as .5000 . . . and not as .4999 . . . . In this way we guarantee that each real number has one and only one decimal representation. Suppose now that we can list all the real numbers, and that they have been listed in a column like the one below (where we use particular numbers for the purpose of illustration).

1st number	13 + .712983 . . .
2nd number	−4 + .913572 . . .
3rd number	0 + .843265 . . .
. . . . .	. . . . .

Since it is impossible actually to write down this infinite list of decimals, our assumption that all the real numbers can be listed in this way means that we assume that we have available some general rule according to which the list is constructed, similar to that used for listing the positive rationals, and that every conceivable real number occurs somewhere in this list. We now demonstrate that this assumption is false by exhibiting a decimal $.a_1a_2a_3 \ldots$ which is constructed in such a way that it is not in the list. We choose a_1 to be 1 unless the first digit after the decimal point of the first number in our list is 1, in which case we choose a_1 to be 2. Clearly, our new decimal will differ from the first number in our list regardless of how we choose its remaining digits. Next, we choose a_2 to be 1 unless the second digit after the decimal point of the second number in our list is 1, in which case we choose a_2 to be 2. Just as above, our new decimal will necessarily differ from the second number in our list. We continue building up the decimal $.a_1a_2a_3 \ldots$ in this way, and since the process can be continued indefinitely, it defines a real number in decimal form (.121 . . . in the case of our illustrative example) which is different from each number in our list. This contradicts our assumption that we can list all the real numbers and completes our proof of the fact that the set R of all real numbers is uncountable.

We have seen (in Problem 6-1) that the set of all rational points on the real line is countable, and we have just proved that the set of *all* points on the real line is uncountable. We conclude at once from this that irrational points on the real line (i.e., irrational numbers) must exist. In fact, it is very easy to see by means of Problem 6-2 that the set of all irrational numbers is uncountably infinite. To vary slightly a striking metaphor coined by E. T. Bell, the rational numbers are spotted along the real line like stars against a black sky, and the dense blackness of the background is the firmament of the irrationals. The reader is probably familiar with a proof of the fact that the square root of 2 is irrational. This proof demonstrates the existence of irrational numbers by exhibiting a specimen. Our remarks, on the other hand, do not show that this or that particular number is irrational; they merely show that such numbers must exist, and moreover must exist in overwhelming abundance.

If the reader supposes that the set of all points on the real line R is uncountable because R is infinitely long, then we can disillusion him by the following argument, which shows that any open interval on R, no matter how short it may be, has precisely as many points as R itself. Let a and b be any two real numbers with $a < b$, and consider the open interval (a,b). Figure 14 shows how to establish a one-to-one correspondence between the points P of (a,b) and the points P' of R: we bend (a,b) into a semicircle; we rest this semicircle tangentially on the

real line R as shown in the figure; and we link P and P' by projecting from its center. If formulas are preferred over geometric reasoning of this kind, we observe that $y = a + (b - a)x$ is a numerical equivalence between real numbers $x \in (0,1)$ and $y \in (a,b)$, and that $z = \tan \pi(x - \frac{1}{2})$ is another numerical equivalence between $(0,1)$ and all of R. It now follows that (a,b) and R are numerically equivalent to one another.

We are now in a position to show that any subset X of the real line R which contains an open interval I is numerically equivalent to R, no matter how complicated the structure of X may be. The proof of this fact is very simple, and it uses only the Schroeder-Bernstein theorem and our above result that I is numerically equivalent to R. The argument can be given in two sentences. Since X is numerically equivalent to itself, it is obviously numerically equivalent to a subset of R; and R is numerically equivalent to a subset of X, namely, to I. It is now a direct consequence of the Schroeder-Bernstein theorem that X and R are numerically equivalent to one another. We point out that all numerical equivalences up to this point have been established by actually exhibiting one-to-one correspondences between the sets concerned. In the present situation, however, it is not feasible to do this, for very little has been assumed about the specific nature of the set X. Without the help of the Schroeder-Bernstein theorem it would be very difficult to prove theorems of this type.

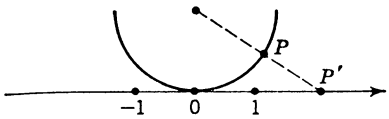

Fig. 14. A one-to-one correspondence between an open interval and the real line.

We give another interesting application of the Schroeder-Bernstein theorem. Consider the coordinate plane R^2 and the subset X of R^2 defined by $X = \{(x,y) : 0 \leq x < 1 \text{ and } 0 \leq y < 1\}$. We show that X is numerically equivalent to the closed-open interval

$$I = \{(x,y) : 0 \leq x < 1 \text{ and } y = 0\}$$

which forms its base (see Fig. 15). Since I is numerically equivalent to a subset of X, namely, to I itself, our conclusion will follow at once from the Schroeder-Bernstein theorem if we can establish a one-to-one mapping of X into I. This we now do. Let (x,y) be an arbitrary point of X. Each of the coordinates x and y has a unique decimal expansion which does not end in an infinite chain of 9's. We form another decimal z from these by alternating their digits; for example, if $x = .327 \ldots$ and $y = .614 \ldots$, then $z = .362174 \ldots$. We now identify z (which cannot end in an infinite chain of 9's) with a point of I. This gives the required one-to-one mapping of X into I and yields the somewhat

startling result that there are no more points inside a square than there are on one of its sides.

In Sec. 6 we introduced the symbol $\aleph_0$ for the number of elements in any countably infinite set. At the beginning of this section we proved that the set R of all real numbers (or of all points on the real line) is uncountably infinite. We now introduce the symbol c (called the *cardinal number of the continuum*) for the number of elements in R. c is the cardinal number of R and of any set which is numerically equivalent to R. In the above three paragraphs we have demonstrated that c is the cardinal number of any open interval, of any subset of R which contains an open interval, and of the subset X of the coordinate plane which is illustrated in Fig. 15. Our list of cardinal numbers has now grown to

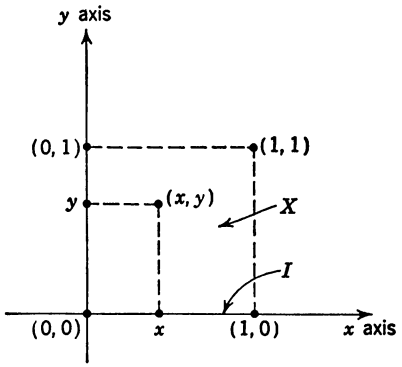

Fig. 15

$$1, 2, 3, \ldots, \aleph_0, c,$$

and they are related to each other by

$$1 < 2 < 3 < \cdots < \aleph_0 < c.$$

At this point we encounter one of the most famous unsolved problems of mathematics. Is there a cardinal number greater than $\aleph_0$ and less than c? No one knows the answer to this question. Cantor himself thought that there is no such number, or in other words, that c is the next infinite cardinal number greater than $\aleph_0$, and his guess has come to be known as *Cantor's continuum hypothesis.* The continuum hypothesis can also be expressed by the assertion that every uncountable set of real numbers has c as its cardinal number.[1]

There is another question which arises naturally at this stage, and this one we are fortunately able to answer. Are there any infinite cardinal numbers greater than c? Yes, there are; for example, the cardinal number of the class of all subsets of R. This answer depends on the following fact: if X is any non-empty set, then the cardinal number of X is less than the cardinal number of the class of all subsets of X.

We prove this statement as follows. In accordance with the definition given in the last paragraph of the previous section, we must show

[1] For further information about the continuum hypothesis, see Wilder [42, p. 125] and Gödel [12].

(1) that there exists a one-to-one mapping of X into the class of all its subsets, and (2) that there does not exist such a mapping of X onto this class. To prove (1), we have only to point to the mapping $x \to \{x\}$, which makes correspond to each element x that set $\{x\}$ which consists of the element x alone. We prove (2) indirectly. Let us assume that there does exist a one-to-one mapping f of X onto the class of all its subsets. We now deduce a contradiction from the assumed existence of such a mapping. Let A be the subset of X defined by $A = \{x : x \notin f(x)\}$. Since our mapping f is onto, there must exist an element a in X such that $f(a) = A$. Where is the element a? *If* a is in A, then by the definition of A we have $a \notin f(a)$, and since $f(a) = A$, $a \notin A$. This is a contradiction, so a cannot belong to A. But if a is not in A, then again by the definition of A we have $a \in f(a)$ or $a \in A$, which is another contradiction. The situation is impossible, so our assumption that such a mapping exists must be false.

This result guarantees that given any cardinal number, there always exists a greater one. If we start with a set $X_1 = \{1\}$ containing one element, then there are two subsets, the empty set $\emptyset$ and the set $\{1\}$ itself. If $X_2 = \{1,2\}$ is a set containing two elements, then there are four subsets: $\emptyset$, $\{1\}$, $\{2\}$, $\{1,2\}$. If $X_3 = \{1, 2, 3\}$ is a set containing three elements, then there are eight subsets: $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1, 2, 3\}$. In general, if X_n is a set with n elements, where n is any finite cardinal number, then X_n has 2^n subsets. If we now take n to be any infinite cardinal number, the above facts suggest that we *define* 2^n to be the number of subsets of any set with n elements. If n is the first infinite cardinal number, namely, $\aleph_0$, then it can be shown that

$$2^{\aleph_0} = c.$$

The simplest proof of this fact depends on the ideas developed in the following paragraph.

Consider the closed-open unit interval $[0,1)$ and a real number x in this set. Our concern is with the meaning of the *decimal*, *binary*, and *ternary expansions* of x. For the sake of clarity, let us take x to be $\frac{1}{4}$. How do we arrive at the decimal expansion of $\frac{1}{4}$? First, we split $[0,1)$ into the 10 closed-open intervals

$$[0,\tfrac{1}{10}), [\tfrac{1}{10},\tfrac{2}{10}), \ldots, [\tfrac{9}{10},1),$$

and we use the 10 digits 0, 1, . . . , 9 to number them in order. Our number $\frac{1}{4}$ belongs to exactly one of these intervals, namely, to $[\frac{2}{10},\frac{3}{10})$. We have labeled this interval with the digit 2, so 2 is the first digit after the decimal point in the decimal expansion of $\frac{1}{4}$:

$$\tfrac{1}{4} = .2 \ldots$$

Next, we split the interval $[2/10, 3/10)$ into the 10 closed-open intervals

$$[2/10, 21/100), [21/100, 22/100), \ldots, [29/100, 3/10),$$

and we use the 10 digits to number these in order. Our number $1/4$ belongs to $[25/100, 26/100)$, which is labeled with the digit 5, so 5 is the second number after the decimal point in the decimal expansion of $1/4$:

$$1/4 = .25 \ldots$$

If we continue this process exactly as we started it, we can obtain the decimal expansion of $1/4$ to as many places as we wish. As a matter of fact, if we do continue, we get 0 at each stage from this point on:

$$1/4 = .25000 \ldots$$

The reader should notice that there is no ambiguity in this system as we have explained it: contrary to customary usage, .24999 . . . is *not* to be regarded as another decimal expansion of $1/4$ which is "equivalent" to .25000 . . . . In this system, each real number x in $[0,1)$ has *one and only one* decimal expansion which cannot end in an infinite chain of 9's. There is nothing magical about the role of the number 10 in the above discussion. If at each stage we split our closed-open interval into two equal closed-open intervals, and if we use the two digits 0 and 1 to number them, we obtain the binary expansion of any real number x in $[0,1)$. The binary expansion of $1/4$ is easily seen to be .01000 . . . . The ternary expansion of x is found similarly: at each stage we split our closed-open interval into three equal closed-open intervals, and we use the three digits 0, 1, and 2 to number them. A moment's thought should convince the reader that the ternary expansion of $1/4$ is .020202 . . . . Just as (in our system) the decimal expansion of a number in $[0,1)$ cannot end in an infinite chain of 9's, so also its binary expansion cannot end in an infinite chain of 1's, and its ternary expansion cannot end in an infinite chain of 2's.

We now use this machinery to give a proof of the fact that

$$2^{\aleph_0} = c.$$

Consider the two sets $N = \{1, 2, 3, \ldots\}$ and $I = [0,1)$, the first with cardinal number $\aleph_0$ and the second with cardinal number c. If $\mathbf{N}$ denotes the class of all subsets of N, then by definition $\mathbf{N}$ has cardinal number $2^{\aleph_0}$. Our proof amounts to showing that there exists a one-to-one correspondence between $\mathbf{N}$ and I. We begin by establishing a one-to-one mapping f of $\mathbf{N}$ into I. If A is a subset of N, then $f(A)$ is that real number x in I whose decimal expansion $x = .d_1d_2d_3 \ldots$ is defined by the condition that d_n is 3 or 5 according as n is or is not in A. Any other two digits can be used here, as long as neither of them is 9. Next, we con-

struct a one-to-one mapping g of I into $\mathbf{N}$. If x is a real number in I, and if $x = .b_1b_2b_3 \ldots$ is its binary expansion (so that each b_n is either 0 or 1), then $g(x)$ is that subset A of N defined by $A = \{n : b_n = 1\}$. We conclude the proof with an appeal to the Schroeder-Bernstein theorem, which guarantees that under these conditions $\mathbf{N}$ and I are numerically equivalent to one another.

If we follow up the hint contained in the fact that $2^{\aleph_0} = c$, and successively form 2^c, 2^{2^c}, and so on, we get a chain of cardinal numbers

$$1 < 2 < 3 < \cdots < \aleph_0 < c < 2^c < 2^{2^c} < \cdots$$

in which there are infinitely many infinite cardinal numbers. Clearly, there is only one kind of countable infinity, symbolized by $\aleph_0$, and beyond this there is an infinite hierarchy of uncountable infinities which are all distinct from one another.

At this point we bring our discussion of these matters to a close. We have barely touched on Cantor's theory and have left entirely to one side, for instance, all questions relating to the addition and multiplication of infinite cardinal numbers and the rules of arithmetic which apply to these operations. We have developed these ideas, not for their own sake, but for the sake of their applications in algebra and topology, and our main purpose throughout the last two sections has been to give the reader some of the necessary insight into countable and uncountable sets and the distinction between them.[1]

Problems

1. Show geometrically that the set of all points in the coordinate plane R^2 is numerically equivalent to the subset X of R^2 illustrated in Fig. 15 and defined by $X = \{(x,y) : 0 \leq x < 1 \text{ and } 0 \leq y < 1\}$, and that therefore R^2 has cardinal number c. [*Hint:* rest an open hemispherical surface (= a hemispherical surface minus its boundary) tangentially on the center of X, project from various points on the line through its center and perpendicular to R^2, and use the Schroeder-Bernstein theorem.]
2. Show that the subset X of R^3 defined by

$$X = \{(x_1, x_2, x_3) : 0 \leq x_i < 1 \text{ for } i = 1, 2, 3\}$$

has cardinal number c.

[1] For the reader who wishes to learn something about the arithmetic of infinite cardinal numbers, we recommend Halmos [16, sec. 24], Kamke [24, chap. 2], Sierpinski [37, chaps. 7–10], or Fraenkel [9, chap. 2].

3. Let n be a positive integer and consider a polynomial equation of the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0,$$

with integral coefficients and $a_n \neq 0$. Such an equation has precisely n complex roots (some of which, of course, may be real). An *algebraic number* is a complex number which is a root of such an equation. The set of all algebraic numbers contains the set of all rational numbers (e.g., $\frac{2}{3}$ is the root of $3x - 2 = 0$) and many other numbers besides (the square root of 2 is a root of $x^2 - 2 = 0$, and $1 + i$ is a root of $x^2 - 2x + 2 = 0$). Complex numbers which are not algebraic are called *transcendental.* The numbers e and π are the best known transcendental numbers, though the fact that they are transcendental is quite difficult to prove (see Niven [33, chap. 9]). Prove that real transcendental numbers exist (*hint:* see Problem 6-5). Prove also that the set of all real transcendental numbers is uncountably infinite.

4. Prove that every infinite set is numerically equivalent to a proper subset of itself (*hint:* see Problem 6-6).

5. Prove that the set of all real functions defined on the closed unit interval has cardinal number 2^c. [*Hint:* there are at least as many such functions as there are *characteristic functions* (i.e., functions whose values are 0 or 1) defined on the closed unit interval.]

8. PARTIALLY ORDERED SETS AND LATTICES

There are two types of relations which often arise in mathematics: order relations and equivalence relations. We touched briefly on order relations in Problem 1-2, and in Section 5 we discussed equivalence relations in some detail. We now return to the topic of order relations and develop those parts of this subject which are necessary for our later work. The reader will find it helpful to keep in mind that a partial order relation (as we define it below) is a generalization of both set inclusion and the order relation on the real line.

Let P be a non-empty set. A *partial order relation* in P is a relation which is symbolized by $\leq$ and assumed to have the following properties:

(1) $x \leq x$ for every x (*reflexivity*);

(2) $x \leq y$ and $y \leq x \Rightarrow x = y$ (*antisymmetry*);

(3) $x \leq y$ and $y \leq z \Rightarrow x \leq z$ (*transitivity*).

We sometimes write $x \leq y$ in the equivalent form $y \geq x$. A non-empty set P in which there is defined a partial order relation is called a *partially*

ordered set. It is clear that any non-empty subset of a partially ordered set is a partially ordered set in its own right.

Partially ordered sets are abundant in all branches of mathematics. Some are simple and easy to grasp, while others are complex and rather inaccessible. We give four examples which are quite different in nature but possess in common the virtues of being both important and easily described.

Example 1. Let P be the set of all positive integers, and let $m \leq n$ mean that m divides n.

Example 2. Let P be the set R of all real numbers, and let $x \leq y$ have its usual meaning (see Problem 1-2).

Example 3. Let $\mathbf{P}$ be the class of all subsets of some universal set U, and let $A \leq B$ mean that A is a subset of B.

Example 4. Let P be the set of all real functions defined on a non-empty set X, and let $f \leq g$ mean that $f(x) \leq g(x)$ for every x.

Two elements x and y in a partially ordered set are called *comparable* if one of them is less than or equal to the other, that is, if either $x \leq y$ or $y \leq x$. The word "partially" in the phrase "partially ordered set" is intended to emphasize that there may be pairs of elements in the set which are not comparable. In Example 1, for instance, the integers 4 and 6 are not comparable, because neither divides the other; and in Example 3, if the universal set U has more than one element, it is always possible to find two subsets of U neither of which is a subset of the other.

Some partial order relations possess a fourth property in addition to the three required by the definition:

(4) any two elements are comparable.

A partial order relation with property (4) is called a *total* (or *linear*) *order relation*, and a partially ordered set whose relation satisfies condition (4) is called a *totally ordered set*, or a *linearly ordered set*, or, most frequently, a *chain*. Example 2 is a chain, as is the subset $\{2, 4, 8, \ldots, 2^n, \ldots\}$ of Example 1.

Let P be a partially ordered set. An element x in P is said to be *maximal* if $y \geq x \Rightarrow y = x$, that is, if no element other than x itself is greater than or equal to x. A maximal element in P is thus an element of P which is not less than or equal to any other element of P. Examples 1, 2, and 4 have no maximal elements. Example 3 has a single maximal element: the set U itself.

Let A be a non-empty subset of a partially ordered set P. An element x in P is called a *lower bound* of A if $x \leq a$ for each $a \in A$; and a lower bound of A is called a *greatest lower bound* of A if it is greater than or

equal to every lower bound of A. Similarly, an element y in P is said to be an *upper bound* of A if $a \leq y$ for every $a \in A$; and a *least upper bound* of A is an upper bound of A which is less than or equal to every upper bound of A. In general, A may have many lower bounds and many upper bounds, but it is easy to prove (see Problem 1) that a greatest lower bound (or least upper bound) is unique if it exists. It is therefore legitimate to speak of *the* greatest lower bound and *the* least upper bound if they exist.

We illustrate these concepts in some of the partially ordered sets mentioned above.

In Example 1, let the subset A consist of the integers 4 and 6. An upper bound of $\{4,6\}$ is any positive integer divisible by both 4 and 6. 12, 24, 36, and so on, are all upper bounds of $\{4,6\}$. 12 is clearly its least upper bound, for it is less than or equal to (i.e., it divides) every upper bound. The greatest lower bound of any pair of integers in this example is their greatest common divisor, and their least upper bound is their least common multiple—both of which are familiar notions from elementary arithmetic.

We now consider Example 2, the real line with its natural order relation. The reader will doubtless recall from his study of calculus that 3 is an upper bound of the set $\{(1 + 1/n)^n : n = 1, 2, 3, \ldots\}$ and that its least upper bound is the fundamental constant $e = 2.7182 \ldots$. As we have stated before, it is a basic property of the real line that every non-empty subset of it which has a lower bound (or upper bound) has a greatest lower bound (or least upper bound). There are several items of standard notation and terminology which must be mentioned in connection with this example. Let A be any non-empty set of real numbers. If A has a lower bound, then its greatest lower bound is usually called its *infimum* and denoted by inf A. Correspondingly, if A has an upper bound, then its least upper bound is called its *supremum* and written sup A. If A happens to be finite, then inf A and sup A both exist and belong to A. In this case, they are often called the *minimum* and *maximum* of A and are denoted by min A and max A. If A consists of two real numbers a_1 and a_2, then min A is the smaller of a_1 and a_2, and max A is the larger.

Finally, consider Example 3, and let $\mathbf{A}$ be any non-empty class of subsets of U. A lower bound of $\mathbf{A}$ is any subset of U which is contained in every set in $\mathbf{A}$, and the greatest lower bound of $\mathbf{A}$ is the intersection of all its sets. Similarly, the least upper bound of $\mathbf{A}$ is the union of all its sets.

One of our main aims in this section is to state *Zorn's lemma*, an exceedingly powerful tool of proof which is almost indispensable in many parts of modern pure mathematics. Zorn's lemma asserts that

if P is a partially ordered set in which every chain has an upper bound, then P possesses a maximal element. It is not possible to prove this in the usual sense of the word. However, it can be shown that Zorn's lemma is logically equivalent to the *axiom of choice*, which states: given any non-empty class of disjoint non-empty sets, a set can be formed which contains precisely one element taken from each set in the given class. The axiom of choice may strike the reader as being intuitively obvious, and in fact, either this axiom itself or some other principle equivalent to

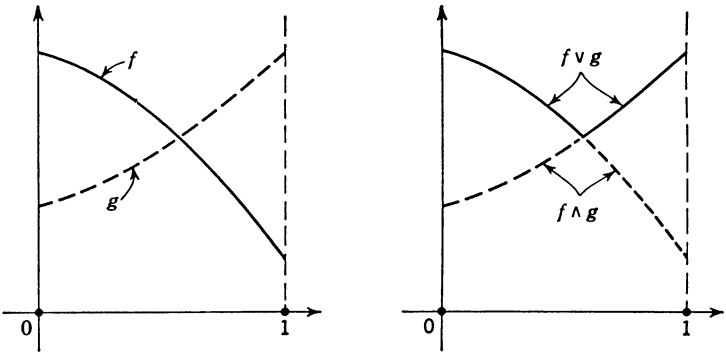

Fig. 16. The geometric meaning of $f \wedge g$ and $f \vee g$.

it is usually postulated in the logic with which we operate. We therefore assume Zorn's lemma as an axiom of logic. Any reader who is interested in these matters is urged to explore them further in the literature.[1]

A *lattice* is a partially ordered set L in which each pair of elements has a greatest lower bound and a least upper bound. If x and y are two elements in L, we denote their greatest lower bound and least upper bound by $x \wedge y$ and $x \vee y$. These notations are analogous to (and are intended to suggest) the notations for the intersection and union of two sets. We pursue this analogy even further, and call $x \wedge y$ and $x \vee y$ the *meet* and *join* of x and y. It is tempting to assume that all properties of intersections and unions in the algebra of sets carry over to lattices, but this is not a valid assumption. Some properties do carry over (see Problem 5), but others, for instance the distributive laws, are false in some lattices.

It is easy to see that all four of our examples are lattices. In Example 1, $m \wedge n$ is the greatest common divisor of m and n, and $m \vee n$ is their least common multiple; and in Example 3, $A \wedge B = A \cap B$ and $A \vee B = A \cup B$. In Example 2, if x and y are any two real numbers, then $x \wedge y$ is min $\{x,y\}$ and $x \vee y$ is max $\{x,y\}$. In Example 4, $f \wedge g$ is

[1] See, for example, Wilder [42, pp. 129–132], Halmos [16, secs. 15–16], Birkhoff [4, p. 42], Sierpinski [37, chap. 6], or Fraenkel and Bar-Hillel [10, p. 44].

the real function defined on X by $(f \wedge g)(x) = \min \{f(x),g(x)\}$, and $f \vee g$ is that defined by $(f \vee g)(x) = \max \{f(x),g(x)\}$. Figure 16 illustrates the geometric meaning of $f \wedge g$ and $f \vee g$ for two real functions f and g defined on the closed unit interval [0,1].

Let L be a lattice. A *sublattice* of L is a non-empty subset L_1 of L with the property that if x and y are in L_1, then $x \wedge y$ and $x \vee y$ are also in L_1. If L is the lattice of all real functions defined on the closed unit interval, and if L_1 is the set of all continuous functions in L, then L_1 is easily seen to be a sublattice of L.

If a lattice has the additional property that every non-empty subset has a greatest lower bound and a least upper bound, then it is called a *complete lattice.* Example 3 is the only complete lattice in our list.

There are many distinct types of lattices, and the theory of these systems has a wide variety of interesting and significant applications (see Birkhoff [4]). We discuss some of these types in our Appendix on Boolean algebras.

Problems

1. Let A be a non-empty subset of a partially ordered set P. Show that A has at most one greatest lower bound and at most one least upper bound.
2. Consider the set $\{1, 2, 3, 4, 5\}$. What elements are maximal if it is ordered as Example 1? If it is ordered as Example 2?
3. Under what circumstances is Example 4 a chain?
4. Give an example of a partially ordered set which is not a lattice.
5. Let L be a lattice. If x, y, and z are elements of L, verify the following: $x \wedge x = x$, $x \vee x = x$, $x \wedge y = y \wedge x$, $x \vee y = y \vee x$,

$$x \wedge (y \wedge z) = (x \wedge y) \wedge z,$$

$x \vee (y \vee z) = (x \vee y) \vee z$, $(x \wedge y) \vee x = x$, $(x \vee y) \wedge x = x$.
6. Let **A** be a class of subsets of some non-empty universal set U. We say that **A** has the *finite intersection property* if every finite subclass of **A** has non-empty intersection. Use Zorn's lemma to prove that if **A** has the finite intersection property, then it is contained in some maximal class **B** with this property (to say that **B** is a *maximal* class with this property is to say that any class which properly contains **B** fails to have this property). (*Hint:* consider the family of all classes which contain **A** and have the finite intersection property, order this family by class inclusion, and show that any chain in the family has an upper bound in the family.)
7. Prove that if X and Y are any two non-empty sets, then there exists a one-to-one mapping of one into the other. (*Hint:* choose an

element x in X and an element y in Y, and establish the obvious one-to-one correspondence between the two single-element sets $\{x\}$ and $\{y\}$; define an *extension* to be a pair of subsets A of X and B of Y such that $\{x\} \subseteq A$ and $\{y\} \subseteq B$, together with a one-to-one correspondence between them under which x and y correspond with one another; order the set of all extensions in the natural way; and apply Zorn's lemma.)

8. Let m and n be any two cardinal numbers (finite or infinite). The statement that m is *less than or equal to* n (written $m \leq n$) is defined to mean the following: if X and Y are sets with m and n elements, then there exists a one-to-one mapping of X into Y. Prove that any non-empty set of cardinal numbers forms a chain when it is ordered in this way. The fact that for any two cardinal numbers one is less than or equal to the other is usually called the *comparability theorem for cardinal numbers.*

9. Let X and Y be non-empty sets, and show that the cardinal number of X is less than or equal to the cardinal number of $Y \Leftrightarrow$ there exists a mapping of Y onto X.

10. Let $\{X_i\}$ be any infinite class of countable sets indexed by the elements i of an index set I, and show that the cardinal number of $\bigcup_i X_i$ is less than or equal to the cardinal number of I. (*Hint:* if I is only countably infinite, this follows from Problem 6-2, and if I is uncountable, Zorn's lemma can be applied to represent it as the union of a disjoint class of countably infinite subsets.)