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.

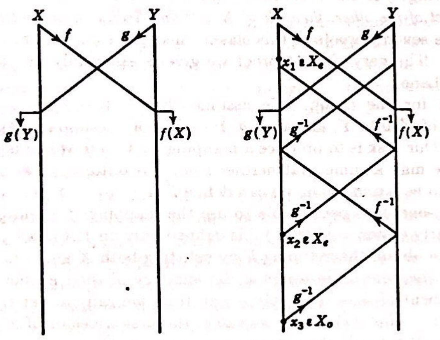

Fig. 12. The proof of the Schroeder-Bernstein theorem.

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:

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

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:

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

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, $1/1 = 1$. Next we list (with increasing numerators) all those for which this sum is $3: 1/2, 2/1 = 2$. Next, all those for which this sum is $4: 1/3, 2/2 = 1, 3/1 = 3$. Next, all those for which this sum is $5: 1/4, 2/3, 3/2, 4/1 = 4$. Next, all those for which this sum is $6: 1/5, 2/4 = 1/2, 3/3 = 1, 4/2 = 2, 5/1 = 5$. 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

$$1,\ 1/2,\ 2,\ 1/3,\ 3,\ 1/4,\ 2/3,\ 3/2,\ 4,\ 1/5,\ 5,\ \ldots$$

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

$$\begin{array}{cccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots \\
\frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \frac{2}{5} & \cdots \\
\frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \frac{3}{5} & \cdots \\
\frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \frac{4}{5} & \cdots \\
\frac{5}{1} & \frac{5}{2} & \frac{5}{3} & \frac{5}{4} & \frac{5}{5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}$$

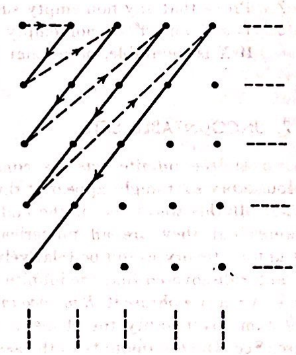

Fig. 13. A listing of the positive rationals.

(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	$13 + .712983 \ldots$
2nd number	$-4 + .913572 \ldots$
3rd number	$0 + .843265 \ldots$
. . . .	. . . .

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.

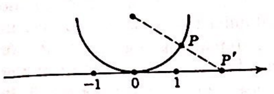

Fig. 14. A one-to-one correspondence between an open interval and the real line.

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 $X = \{(x,y): 0 \le x < 1 \text{ and } 0 \le y < 1\}$. We show that X is numerically equivalent to the closed-open interval

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

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

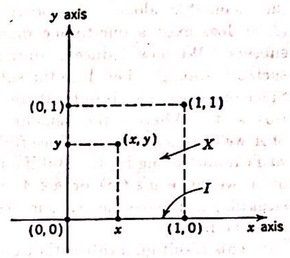

Fig. 15

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

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 $[\frac{2}{10},\frac{3}{10})$ into the 10 closed-open intervals

$$[\tfrac{2}{10},\tfrac{21}{100}),\ [\tfrac{21}{100},\tfrac{22}{100}),\ \ldots,\ [\tfrac{29}{100},\tfrac{3}{10}),$$

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

$$\tfrac{1}{4} = .25\ \ldots$$

If we continue this process exactly as we started it, we can obtain the decimal expansion of $\frac{1}{4}$ 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:

$$\tfrac{1}{4} = .25000\ \ldots$$

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 $\frac{1}{4}$ 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 $\frac{1}{4}$ 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 $\frac{1}{4}$ 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 N denotes the class of all subsets of N, then by definition N has cardinal number $2^{\aleph_0}$. Our proof amounts to showing that there exists a one-to-one correspondence between N and I. We begin by establishing a one-to-one mapping f of 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 construct a one-to-one mapping g of I into N. If x is a real number in I, and if $x = .b_1b_2b_3 \cdots$ 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 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

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

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., $2/3$ 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 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 A be any non-empty class of subsets of U. A lower bound of A is any subset of U which is contained in every set in A, and the greatest lower bound of A is the intersection of all its sets. Similarly, the least upper bound of 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 the following: given any non-empty class of 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

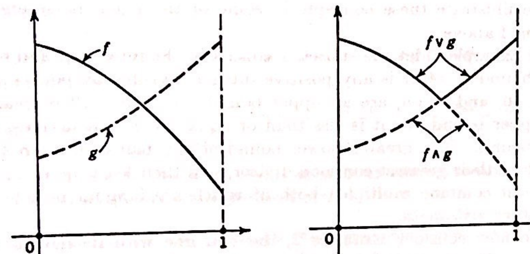

Fig. 16. The geometric meaning of $f \wedge g$ and $f \vee g$.

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 $\cup_i X_i$ 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.)

CHAPTER TWO

Metric Spaces

Classical analysis can be described as that part of mathematics which begins with calculus and, in essentially the same spirit, develops similar subject matter much further in many directions. It is a great nation in the world of mathematics, with many provinces, a few of which are ordinary and partial differential equations, infinite series (especially power series and Fourier series), and analytic functions of a complex variable. Each of these has experienced enormous growth over a long history, and each is rich enough in content to merit a lifetime of study.

In the course of its development, classical analysis became so complex and varied that even an expert could find his way around in it only with difficulty. Under these circumstances, some mathematicians became interested in trying to uncover the fundamental principles on which all analysis rests. This movement had associated with it many of the great names in mathematics of the last century: Riemann, Weierstrass, Cantor, Lebesgue, Hilbert, Riesz, and others. It played a large part in the rise to prominence of topology, modern algebra, and the theory of measure and integration; and when these new ideas began to percolate back through classical analysis, the brew which resulted was modern analysis.

As modern analysis developed in the hands of its creators, many a major theorem was given a simpler proof in a more general setting, in an effort to lay bare its inner meaning. Much thought was devoted to analyzing the texture of the real and complex number systems, which are the context of analysis. It was hoped—and these hopes were well founded—that analysis could be clarified and simplified, and that stripping away