2. Equivalence of Sets. The Power of a Set

2.1. Finite and infinite sets. The set of all vertices of a given polyhedron, the set of all prime numbers less than a given number, and the set of all

residents of New York City (at a given time) have a certain property in common, namely, each set has a definite number of elements which can be found in principle, if not in practice. Accordingly, these sets are all said to be *finite*. Clearly, we can be sure that a set is finite without knowing the number of elements in it. On the other hand, the set of all positive integers, the set of all points on the line, the set of all circles in the plane, and the set of all polynomials with rational coefficients have a different property in common, namely, if we remove one element from each set, then remove two elements, three elements, and so on, there will still be elements left in the set at each stage. Accordingly, sets of this kind are said to be *infinite*.

Given two finite sets, we can always decide whether or not they have the same number of elements, and if not, we can always determine which set has more elements than the other. It is natural to ask whether the same is true of infinite sets. In other words, does it make sense to ask, for example, whether there are more circles in the plane than rational points on the line, or more functions defined in the interval $[0, 1]$ than lines in space? As will soon be apparent, questions of this kind can indeed be answered.

To compare two finite sets A and B, we can count the number of elements in each set and then compare the two numbers, but alternatively, we can try to establish a *one-to-one correspondence* between (the elements of) A and B, i.e., a correspondence such that each element in A corresponds to one and only one element in B and vice verse. It is clear that a one-to-one correspondence between two finite sets can be set up if and only if the two sets have the same number of elements. For example, to ascertain whether or not the number of students in an assembly is the same as the number of seats in the auditorium, there is no need to count the number of students and the number of seats. We need merely observe whether or not there are empty seats or students with no place to sit down. If the students can all be seated with no empty seats left, i.e., if there is a one-to-one correspondence between the set of students and the set of seats, then these two sets obviously have the same number of elements. The important point here is that the first method (counting elements) works only for finite sets, while the second method (setting up a one-to-one correspondence) works for infinite sets as well as for finite sets.

2.2. Countable sets. The simplest infinite set is the set Z_+ of all positive integers. An infinite set is called *countable* if its elements can be put in one-to-one correspondence with those of Z_+. In other words, a countable set is a set whose elements can be numbered $a_1, a_2, \ldots, a_n, \ldots$. By an *uncountable* set we mean, of course, an infinite set which is not countable.

We now give some examples of countable sets:

Example 1. The set Z of all integers, positive, negative or zero, is countable. In fact, we can set up the following one-to-one correspondence

between Z and the set Z_+ of all positive integers:

$$\begin{matrix} 0, & -1, & 1, & -2, & 2, \ldots \\ 1, & 2, & 3, & 4, & 5, \ldots \end{matrix}$$

More explicitly, we associate the nonnegative integer $n \geq 0$ with the odd number $2n + 1$, and the negative integer $n < 0$ with the even number $2\,|n|$, i.e.,

$$n \leftrightarrow 2n + 1 \quad \text{if } n \geq 0,$$
$$n \leftrightarrow 2\,|n| \quad \text{if } n < 0$$

(the symbol $\leftrightarrow$ denotes a one-to-one correspondence).

Example 2. The set of all positive even numbers is countable, as shown by the obvious correspondence $n \leftrightarrow 2n$.

Example 3. The set $2, 4, 8, \ldots, 2^n, \ldots$ of powers of 2 is countable, as shown by the obvious correspondence $n \leftrightarrow 2^n$.

Example 4. The set Q of all rational numbers is countable. To see this, we first note that every rational number α can be written as a fraction p/q, $q > 0$ in lowest terms with a positive denominator. Call the sum $|p| + q$ the "height" of the rational number α. For example,

$$\frac{0}{1} = 0$$

is the only rational number of height 0,

$$\frac{-1}{1}, \quad \frac{1}{1}$$

are the only rational numbers of height 2,

$$\frac{-2}{1}, \quad \frac{-1}{2}, \quad \frac{1}{2}, \quad \frac{2}{1}$$

are the only rational numbers of height 3, and so on. We can now arrange all rational numbers in order of increasing height (with the numerators increasing in each set of rational numbers of the same height). In other words, we first count the rational numbers of height 1, then those of height 2 (suitably arranged), those of height 3, and so on. In this way, we assign every rational number a unique positive integer, i.e., we set up a one-to-one correspondence between the set Q of all rational numbers and the set Z_+ of all positive integers.

Next we prove some elementary theorems involving countable sets:

THEOREM 1. *Every subset of a countable set is countable.*

Proof. Let A be countable, with elements $a_1, a_2, \ldots$, and let B be a subset of A. Among the elements $a_1, a_2, \ldots$, let $a_{n_1}, a_{n_2}, \ldots$ be those in the set B. If the set of numbers $n_1, n_2, \ldots$ has a largest number, then B is finite. Otherwise B is countable (consider the correspondence $i \leftrightarrow a_{n_i}$). ∎

THEOREM 2. *The union of a finite or countable number of countable sets $A_1, A_2, \ldots$ is itself countable.*

Proof. We can assume that no two of the sets $A_1, A_2, \ldots$ have elements in common, since otherwise we could consider the sets

$$A_1,\ A_2 - A_1,\ A_3 - (A_1 \cup A_2), \ldots$$

instead, which are countable by Theorem 1 and have the same union as the original sets. Suppose we write the elements of $A_1, A_2, \ldots$ in the form of an infinite table

$$\begin{array}{llll} a_{11} & a_{12} & a_{13} & a_{14} \cdots \\ a_{21} & a_{22} & a_{23} & a_{24} \cdots \\ a_{31} & a_{32} & a_{33} & a_{34} \cdots \\ a_{41} & a_{42} & a_{43} & a_{44} \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array} \tag{1}$$

where the elements of the set A_1 appear in the first row, the elements of the set A_2 appear in the second row, and so on. We now count all the elements in (1) "diagonally," i.e., first we choose a_{11}, then a_{12}, then a_{21}, and so on, moving in the way shown in the following table:[5]

$$\begin{array}{ccccccc} a_{11} & \to & a_{12} & & a_{13} & \to & a_{14} \cdots \\ & \swarrow & & \nearrow & & \swarrow & \\ a_{21} & & a_{22} & & a_{23} & & a_{24} \cdots \\ \downarrow & \nearrow & & \swarrow & & & \\ a_{31} & & a_{32} & & a_{33} & & a_{34} \cdots \\ & \swarrow & & & & & \\ a_{41} & & a_{42} & & a_{43} & & a_{44} \cdots \\ \cdots & & \cdots & & \cdots & & \cdots \end{array} \tag{2}$$

It is clear that this procedure associates a unique number to each element in each of the sets $A_1, A_2, \ldots$, thereby establishing a one-to-one correspondence between the union of the sets $A_1, A_2, \ldots$ and the set Z_+ of all positive integers. ∎

THEOREM 3. *Every infinite set has a countable subset.*

[5] Discuss the obvious modifications of (1) and (2) in the case of only a finite number of sets $A_1, A_2, \ldots$.

Proof. Let M be an infinite set and a_1 any element of M. Being infinite, M contains an element a_2 distinct from a_1, an element a_3 distinct from both a_1 and a_2, and so on. Continuing this process (which can never terminate due to a "shortage" of elements, since M is infinite), we get a countable subset

$$A = \{a_1, a_2, \ldots, a_n, \ldots\}$$

of the set M. ∎

Remark. Theorem 3 shows that countable sets are the "smallest" infinite sets. The question of whether there exist uncountable (infinite) sets will be considered below.

2.3. Equivalence of sets. We arrived at the notion of a countable set M by considering one-to-one correspondences between M and the set Z_+ of all positive integers. More generally, we can consider one-to-one correspondences between any two sets M and N:

DEFINITION. *Two sets M and N are said to be* **equivalent** *(written $M \sim N$) if there is a one-to-one correspondence between the elements of M and the elements of N.*

The concept of equivalence[6] is applicable to both finite and infinite sets. Two finite sets are equivalent if and only if they have the same number of elements. We can now define a countable set as a set equivalent to the set Z_+ of all positive integers. It is clear that two sets which are equivalent to a third set are equivalent to each other, and in particular that any two countable sets are equivalent.

Example 1. The sets of points in any two closed intervals $[a, b]$ and $[c, d]$ are equivalent, and Figure 5 shows how to set up a one-to-one correspondence between them. Here two points p and q correspond to each other if and only if they lie on the same ray emanating from the point O in which the extensions of the line segments ac and bd intersect.

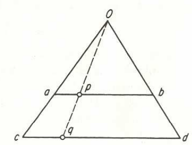

FIGURE 5

Example 2. The set of all points z in the complex plane is equivalent to the set of all

[6] Not to be confused with our previous use of the word in the phrase "equivalence relation." However, note that set equivalence is an equivalence relation in the sense of Sec. 1.4, being obviously reflexive, symmetric and transitive. Hence any family of sets can be partitioned into classes of equivalent sets.

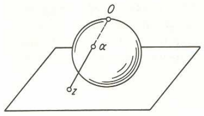

FIGURE 6

points α on a sphere. In fact, a one-to-one correspondence $z \leftrightarrow \alpha$ between the points of the two sets can be established by using stereographic projection, as shown in Figure 6 (O is the north pole of the sphere).

Example 3. The set of all points x in the open unit interval $(0, 1)$ is equivalent to the set of all points y on the whole real line. For example, the formula

$$y = \frac{1}{\pi} \arctan x + \frac{1}{2}$$

establishes a one-to-one correspondence between these two sets.

The last example and the examples in Sec. 2.2 show that an infinite set is sometimes equivalent to one of its proper subsets. For example, there are "as many" positive integers as integers of arbitrary sign, there are "as many" points in the interval $(0, 1)$ as on the whole real line, and so on. This fact is characteristic of all infinite sets (and can be used to define such sets), as shown by

THEOREM 4. *Every infinite set is equivalent to one of its proper subsets.*

Proof. According to Theorem 3, any infinite set M contains a countable subset. Let this subset be

$$A = \{a_1, a_2, \ldots, a_n, \ldots\},$$

and partition A into two countable subsets

$$A_1 = \{a_1, a_3, a_5, \ldots\}, \qquad A_2 = \{a_2, a_4, a_6, \ldots\}.$$

Obviously, we can establish a one-to-one correspondence between the countable sets A and A_1 (merely let $a_n \leftrightarrow a_{2n-1}$). This correspondence can be extended to a one-to-one correspondence between the sets $A \cup (M - A) = M$ and $A_1 \cup (M - A) = M - A_2$ by simply assigning x itself to each element $x \in M - A$. But $M - A_2$ is a proper subset of M. ▮

2.4. Uncountability of the real numbers. Several examples of countable sets were given in Sec. 2.2, and many more examples of such sets could be given. In fact, according to Theorem 2, the union of a finite or countable number of countable sets is itself countable. It is now natural to ask whether there exist infinite sets which are uncountable. The existence of such sets is shown by

THEOREM 5. *The set of real numbers in the closed unit interval* $[0, 1]$ *is uncountable.*

Proof. Suppose we have somehow managed to count some or all of the real numbers in $[0, 1]$, arranging them in a list

$$\begin{array}{l} \alpha_1 = 0.a_{11}a_{12}\ldots a_{1n}\ldots, \\ \alpha_2 = 0.a_{21}a_{22}\ldots a_{2n}\ldots, \\ \ldots\ldots\ldots\ldots\ldots \\ \alpha_n = 0.a_{n1}a_{n2}\ldots a_{nn}\ldots, \\ \ldots\ldots\ldots\ldots\ldots \end{array} \qquad (3)$$

where a_{ik} is the kth digit in the decimal expansion of the number α_i. Consider the decimal

$$\beta = 0.b_1b_2\ldots b_n\ldots \qquad (4)$$

constructed as follows: For b_1 choose any digit (from 0 to 9) different from a_{11}, for b_2 any digit different from a_{22}, and so on, and in general for b_n any digit different from a_{nn}. Then the decimal (4) cannot coincide with any decimal in the list (3). In fact, β differs from α_1 in at least the first digit, from α_2 in at least the second digit, and so on, since in general $b_n \neq a_{nn}$ for all n. Thus no list of real numbers in the interval $[0, 1]$ can include all the real numbers in $[0, 1]$.

The above argument must be refined slightly since certain numbers, namely those of the form $p/10^q$, can be written as decimals in two ways, either with an infinite run of zeros or an infinite run of nines. For example,

$$\tfrac{1}{2} = \tfrac{5}{10} = 0.5000\ldots = 0.4999\ldots,$$

so that the fact that two decimals are distinct does not necessarily mean that they represent distinct real numbers. However, this difficulty disappears if in constructing β, we require that β contain neither zeros nor nines, for example by setting $b_n = 2$ if $a_{nn} = 1$ and $b_n = 1$ if $a_{nn} \neq 1$. ▮

Thus the set $[0, 1]$ is uncountable. Other examples of uncountable sets equivalent to $[0, 1]$ are

1) The set of points in any closed interval $[a, b]$;
2) The set of points on the real line;
3) The set of points in any open interval (a, b);
4) The set of all points in the plane or in space;
5) The set of all points on a sphere or inside a sphere;
6) The set of all lines in the plane;
7) The set of all *continuous* real functions of one or several variables.

The fact that the sets 1) and 2) are equivalent to [0, 1] is proved as in Examples 1 and 3, pp. 13 and 14, while the fact that the sets 3)–7) are equivalent to [0, 1] is best proved *indirectly* (cf. Problems 7 and 9).

2.5. The power of a set. Given any two sets M and N, suppose M and N are equivalent. Then M and N are said to have the same *power*. Roughly speaking, "power" is something shared by equivalent sets. If M and N are finite, then M and N have the same number of elements, and the concept of the power of a set reduces to the usual notion of the number of elements in a set. The power of the set Z_+ of all positive integers, and hence the power of any countable set, is denoted by the symbol $\aleph_0$, read "aleph null." A set equivalent to the set of real numbers in the interval [0, 1], and hence to the set of *all* real numbers, is said to have the power of the *continuum*, denoted by c (or often by $\aleph$).

For the powers of finite sets, i.e., for the positive integers, we have the notions of "greater than" and "less than," as well as the notion of equality. We now show how these concepts are extended to the case of infinite sets.

Let A and B be any two sets, with powers $m(A)$ and $m(B)$, respectively. If A is equivalent to B, then $m(A) = m(B)$ by definition. If A is equivalent to a subset of B and if no subset of A is equivalent to B, then, by analogy with the finite case, it is natural to regard $m(A)$ as less than $m(B)$ or $m(B)$ as greater than $m(A)$. Logically, however, there are two further possibilities:

a) B has a subset equivalent to A, and A has a subset equivalent to B;
b) A and B are not equivalent, and neither has a subset equivalent to the other.

In case a), A and B are equivalent and hence have the same power, as shown by the Cantor-Bernstein theorem (Theorem 7 below). Case b) would obviously show the existence of powers that cannot be compared, but it follows from the well-ordering theorem (see Sec. 3.7) that this case is actually impossible. Therefore, taking both of these theorems on faith, we see that any two sets A and B either have the same power or else satisfy one of the relations $m(A) < m(B)$ or $m(A) > m(B)$. For example, it is clear that $\aleph_0 < c$ (why?).

Remark. The very deep problem of the existence of powers between $\aleph_0$ and c is touched upon in Sec. 3.9. As a rule, however, the infinite sets encountered in analysis are either countable or else have the power of the continuum.

We have already noted that countable sets are the "smallest" infinite sets. It has also been shown that there are infinite sets of power greater than that of a countable set, namely sets with the power of the continuum. It is natural to ask whether there are infinite sets of power greater than that of the continuum or, more generally, whether there is a "largest" power. These questions are answered by

THEOREM 6. *Given any set M, let $\mathscr{M}$ be the set whose elements are all possible subsets of M. Then the power of $\mathscr{M}$ is greater than the power of the original set M.*

Proof. Clearly, the power μ of the set $\mathscr{M}$ cannot be less than the power m of the original set M, since the "single-element subsets" (or "singletons") of M form a subset of $\mathscr{M}$ equivalent to M. Thus we need only show that m and μ do not coincide. Suppose a one-to-one correspondence

$$a \leftrightarrow A, \qquad b \leftrightarrow B, \ldots$$

has been established between the elements $a, b, \ldots$ of M and certain elements $A, B, \ldots$ of $\mathscr{M}$ (i.e., certain subsets of M). Then $A, B, \ldots$ do not exhaust all the elements of $\mathscr{M}$, i.e., all the subsets of M. To see this, let X be the set of elements of M which do not belong to their "associated subsets." More exactly, if $a \leftrightarrow A$ we assign a to X if $a \notin A$, but not if $a \in A$. Clearly, X is a subset of M and hence an element of $\mathscr{M}$. Suppose there is an element $x \in M$ such that $x \leftrightarrow X$, and consider whether or not x belongs to X. Suppose $x \notin X$. Then $x \in X$, since, by definition, X contains *every* element not contained in its associated subset. On the other hand, suppose $x \notin X$. Then $x \in X$, since X consists precisely of those elements which do not belong to their associated subsets. In any event, the element x corresponding to the subset X must simultaneously belong to X and not belong to X. But this is impossible! It follows that there is no such element x. Therefore no one-to-one correspondence can be established between the sets M and $\mathscr{M}$, i.e., $m \neq \mu$. ∎

Thus, given any set M, there is a set $\mathscr{M}$ of larger power, a set $\mathscr{M}^*$ of still larger power, and so on indefinitely. In particular, there is no set of "largest" power.

2.6. The Cantor-Bernstein theorem. Next we prove an important theorem already used in the preceding section:

THEOREM 7 (***Cantor-Bernstein***). *Given any two sets A and B, suppose A contains a subset A_1 equivalent to B, while B contains a subset B_1 equivalent to A. Then A and B are equivalent.*

Proof. By hypothesis, there is a one-to-one function f mapping A into B_1 and a one-to-one function g mapping B into A_1:

$$f(A) = B_1 \subset B, \qquad g(B) = A_1 \subset A.$$

Therefore

$$A_2 = gf(A) = g(f(A)) = g(B_1)$$

is a subset of A_1 equivalent to all of A. Similarly,

$$B_2 = fg(B) = f(g(B)) = f(A_1)$$

is a subset of B_1 equivalent to B. Let A_3 be the subset of A into which the mapping gf carries the set A_1, and let A_4 be the subset of A into which gf carries A_2. More generally, let A_{k+2} be the set into which A_k ($k = 1, 2, \ldots$) is carried by gf. Then clearly

$$A \supset A_1 \supset A_2 \supset \cdots \supset A_k \supset A_{k+1} \supset \cdots$$

Setting

$$D = \bigcap_{k=1}^{\infty} A_k,$$

we can represent A as the following union of pairwise disjoint sets:

$$A = (A - A_1) \cup (A_1 - A_2) \cup (A_2 - A_3) \cup \cdots \cup (A_k - A_{k+1}) \cup \cdots \cup D. \quad (5)$$

Similarly, we can write A_1 in the form

$$A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \cdots \cup (A_k - A_{k+1}) \cup \cdots \cup D. \quad (6)$$

Clearly, (5) and (6) can be rewritten as

$$A = D \cup M \cup N, \quad (5')$$

$$A_1 = D \cup M \cup N_1, \quad (6')$$

where

$$M = (A_1 - A_2) \cup (A_3 - A_4) \cup \cdots,$$
$$N = (A - A_1) \cup (A_2 - A_3) \cup \cdots,$$
$$N_1 = (A_2 - A_3) \cup (A_4 - A_5) \cup \cdots.$$

But $A - A_1$ is equivalent to $A_2 - A_3$ (the former is carried into the latter by the one-to-one function gf), $A_2 - A_3$ is equivalent to $A_4 - A_5$, and so on. Therefore N is equivalent to N_1. It follows from the representations (5′) and (6′) that a one-to-one correspondence can be set up between the sets A and A_1. But A_1 is equivalent to B, by hypothesis. Therefore A is equivalent to B. ∎

Remark. Here we can even "afford the unnecessary luxury" of explicitly writing down a one-to-one function carrying A into B, i.e.,

$$\varphi(a) = \begin{cases} g^{-1}(a) & \text{if } a \in D \cup M, \\ f(a) & \text{if } a \in D \cup N \end{cases}$$

(see Figure 7).

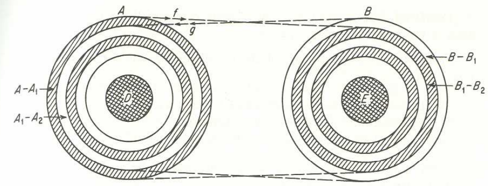

FIGURE 7

Problem 1. Prove that a set with an uncountable subset is itself uncountable.

Problem 2. Let M be any infinite set and A any countable set. Prove that $M \sim M \cup A$.

Problem 3. Prove that each of the following sets is countable:

a) The set of all numbers with two distinct decimal expansions (like 0.5000 . . . and 0.4999 . . .);
b) The set of all rational points in the plane (i.e., points with rational coordinates);
c) The set of all rational intervals (i.e., intervals with rational end points);
d) The set of all polynomials with rational coefficients.

Problem 4. A number α is called *algebraic* if it is a root of a polynomial equation with rational coefficients. Prove that the set of all algebraic numbers is countable.

Problem 5. Prove the existence of uncountably many *transcendental* numbers, i.e., numbers which are not algebraic.

Hint. Use Theorems 2 and 5.

Problem 6. Prove that the set of all real functions (more generally, functions taking values in a set containing at least two elements) defined on a set M is of power greater than the power of M. In particular, prove that the power of the set of *all* real functions (continuous and discontinuous) defined in the interval $[0, 1]$ is greater than c.

Hint. Use the fact that the set of all characteristic functions (i.e., functions taking only the values 0 and 1) on M is equivalent to the set of all subsets of M.

Problem 7. Give an indirect proof of the equivalence of the closed interval $[a, b]$, the open interval (a, b) and the half-open interval $[a, b)$ or $(a, b]$.

Hint. Use Theorem 7.

Filename: reference_2
Directory: C:\Documents and Settings\Dong Won Woo.D62TQS71.001\Desktop\real\equivalence_of_sets_&_the_power_of_a_set
Template: C:\Documents and Settings\Dong Won Woo.D62TQS71.001\Application Data\Microsoft\Templates\Normal.dot
Title:
Subject:
Author: dww
Keywords:
Comments:
Creation Date: 2/21/2009 1:30:00 PM
Change Number: 2
Last Saved On: 2/22/2009 7:37:00 AM
Last Saved By: dww
Total Editing Time: 7 Minutes
Last Printed On: 2/22/2009 7:37:00 AM
As of Last Complete Printing
Number of Pages: 5
Number of Words: 2 (approx.)
Number of Characters: 18 (approx.)