From: "Principles of Mathematical Analysis"
by Walter Rudin
McGraw Hill Pub. Co. (1953)

CHAPTER 2

ELEMENTS OF SET THEORY

Finite, Countable, and Uncountable Sets

We begin this section with a definition of the function concept.

2.1. Definition. Consider two sets A and B, whose elements may be any objects whatsoever, and suppose that with each element x of A there is associated, in some manner, an element of B, which we denote by $f(x)$. Then f is said to be a function from A to B (or, a mapping of A into B). The set A is called the domain of definition of f (we also say, f is defined on A), and the elements $f(x)$ are called the values of f. The set of all values of f is called the range of f.

2.2. Definition. We say that two sets A and B can be put into 1–1 (one-to-one) correspondence if there exists a function f from A to B with the following properties.

(*a*) If $x_1 \in A$, $x_2 \in A$, and $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$.

(*b*) For every $y \in B$ there is an $x \in A$ such that $y = f(x)$.

(The notation $x_1 \neq x_2$ means that x_1 and x_2 are distinct elements; otherwise we write $x_1 = x_2$.)

2.3. Definition. If two sets A and B can be put into 1–1 correspondence, we say that A is equivalent to B (or, A and B are equipotent), and write $A \sim B$. The relation $\sim$ just defined clearly has the following properties:

It is reflexive:	$A \sim A$.
It is symmetric:	if $A \sim B$, then $B \sim A$.
It is transitive:	if $A \sim B$ and $B \sim C$, then $A \sim C$.

Any relation with these three properties is called an equivalence relation.

2.4. Definition. For any positive integer n, let J_n be the set whose elements are the integers, $1, 2, \ldots, n$; let J be the set consisting of all positive integers. For any set A, we say:

(*a*) A is finite if $A \sim J_n$ for some n (the vacuous set is also considered to be finite).

(*b*) A is infinite if A is not finite.

(*c*) A is countable if $A \sim J$.

(*d*) A is uncountable if A is neither finite nor countable.

(*e*) A is at most countable if A is finite or countable.

Countable sets are sometimes called enumerable, or denumerable.

For two finite sets A and B, we evidently have $A \sim B$ if and only if A and B contain the same number of elements. For infinite sets, however, the idea of "having the same number of elements" becomes quite vague, whereas the notion of 1–1 correspondence retains its clarity.

2.5. Example. Let A be the set of all integers. Then A is countable. For, consider the following arrangement of the sets A and J:

$$\begin{array}{ll} A: & 0,\ 1,\ -1,\ 2,\ -2,\ 3,\ -3,\ \ldots \\ J: & 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ \ \ \ \ \ldots \end{array}$$

We can, in this example, even give an explicit formula for the function f from J to A which sets up the 1–1 correspondence:

$$f(n) = \begin{cases} \dfrac{n}{2} & (n \text{ even}), \\ -\dfrac{n-1}{2} & (n \text{ odd}). \end{cases}$$

2.6. Definition. We say that A is a subset of B, and write $A \subset B$ (or $B \supset A$) if every element of A is an element of B. If, in addition, there is an element of B which is not in A, then A is said to be a proper subset of B.

In particular, the vacuous set is a subset of every set, and $A \subset A$ for every set A.

If $A \subset B$ and $B \subset A$, we write $A = B$.

2.7. Remark. The fact that the vacuous set is a subset of every set is based on a point of logic which often causes difficulty to beginners.

By Definition 2.6, it is clear that if A is *not* a subset of B, the following statement must be true: "There is an element x such that $x \in A$ and $x \notin B$." But if A is vacuous, there is no x such that $x \in A$, and the above statement is false.

Similar arguments apply whenever we wish to verify that certain conditions are satisfied by the vacuous set.

2.8. Remark. A finite set cannot be equivalent to one of its proper subsets. That this is, however, possible for infinite sets, is shown by Example 2.5, in which J is a proper subset of A.

In fact, we could replace Definition 2.4(b) by the statement: A is infinite if A is equivalent to one of its proper subsets.

2.9. Definition. By a sequence we mean the values of a function f defined on the set J of all positive integers. If $f(n) = x_n$, for $n \in J$, we use the notation $\{x_n\}$ to denote the sequence whose elements are x_1, x_2, x_3, ...

Note that these elements need not be distinct.

Since every countable set is the range of a 1–1 function defined on J, we can always think of a countable set as a sequence of distinct elements.

2.10. Theorem. *Every infinite subset of a countable set A is countable.*

Proof: Suppose $E \subset A$, and E is infinite. Arrange the elements x of A in a sequence $\{x_n\}$ of distinct elements. Construct a sequence $\{n_k\}$ as follows:

Let n_1 be the smallest positive integer such that $x_{n_1} \varepsilon E$. Having chosen $n_1, \ldots, n_{k-1}$ ($k = 2,3,4, \ldots$), let n_k be the smallest integer greater than n_{k-1} such that $x_{n_k} \varepsilon E$.

Putting $f(k) = x_{n_k}$ ($k = 1,2,3, \ldots$), we obtain a 1–1 correspondence between E and J.

The theorem shows that, roughly speaking, countable sets represent the "smallest" infinity: no uncountable set can be a subset of a countable set.

2.11. Definition. Let A be any set, and suppose that with each element α of A there is associated a set which we denote by E_α. The elements of E_α may or may not be elements of A.

The set whose elements are the sets E_α will be denoted by $\{E_\alpha\}$. Instead of speaking of sets of sets, we shall sometimes speak of a collection of sets, or a family of sets.

The union of the sets E_α is defined to be the set S such that $x \varepsilon S$ if and only if $x \varepsilon E_\alpha$ for at least one $\alpha \varepsilon A$. We use the notation

$$S = \bigcup_{\alpha \varepsilon A} E_\alpha. \tag{1}$$

If A consists of the integers $1,2, \ldots, n$, one usually writes

$$S = \bigcup_{m=1}^{n} E_m \tag{2}$$

or

$$S = E_1 \cup E_2 \cup \cdots \cup E_n. \tag{3}$$

If A is the set of all positive integers, the usual notation is

$$S = \bigcup_{m=1}^{\infty} E_m. \tag{4}$$

The symbol ∞ in (4) merely indicates that the union of a *countable* collection of sets is taken, and should not be confused with the symbols $+\infty$, $-\infty$, introduced in Definition 1.39.

The intersection of the sets E_α is defined to be the set P such that $x \varepsilon P$ if and only if $x \varepsilon E_\alpha$ for every $\alpha \varepsilon A$. We use the notation

$$P = \bigcap_{\alpha \varepsilon A} E_\alpha, \tag{5}$$

or

(6) $$P = \bigcap_{m=1}^{n} E_m = E_1 \cap E_2 \cap \cdots \cap E_n,$$

or

(7) $$P = \bigcap_{m=1}^{\infty} E_m,$$

as for unions.

2.12. Examples. (*a*) Suppose E_1 consists of 1, 2, 3 and E_2 consists of 2, 3, 4. Then $E_1 \cup E_2$ consists of 1, 2, 3, 4, whereas $E_1 \cap E_2$ consists of 2, 3.

(*b*) Let A be the set of real numbers x such that $0 < x < 1$. For every $x \in A$, let E_x be the set of real numbers y such that $0 < y < x$. Then

(i) $$E_x \subset E_z \text{ if and only if } 0 < x < z < 1;$$

(ii) $$\bigcup_{x \in A} E_x = E_1;$$

(iii) $$\bigcap_{x \in A} E_x \text{ is vacuous;}$$

(i) and (ii) are clear. To prove (iii), we note that for every $y > 0$, $y \notin E_x$ if $x < y$. Hence $y \notin \bigcap_{x \in A} E_x$.

2.13. Remarks. Many properties of unions and intersections are quite similar to those of sums and products; in fact, the words sum and product are frequently used in this connection, and the symbols Σ and Π are written in place of $\bigcup$ and $\bigcap$.

The commutative and associative laws are trivial:

(8) $$A \cup B = B \cup A; \qquad A \cap B = B \cap A.$$

(9) $$(A \cup B) \cup C = A \cup (B \cup C); \qquad (A \cap B) \cap C = A \cap (B \cap C).$$

Thus the omission of parentheses in (3) and (6) is justified.

The distributive law also holds:

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

To prove this, let the left and right members of (10) be denoted by E and F, respectively.

Suppose $x \in E$. Then $x \in A$ and $x \in B \cup C$, that is, $x \in B$ or $x \in C$ (possibly both). Hence $x \in A \cap B$ or $x \in A \cap C$, so that $x \in F$. Thus $E \subset F$.

Next, suppose $x \in F$. Then $x \in A \cap B$ or $x \in A \cap C$. That is, $x \in A$, and $x \in B \cup C$. Hence $x \in A \cap (B \cup C)$, so that $F \subset E$.

It follows that $E = F$.

We list a few more relations which are easily verified:

$$A \subset A \cup B, \tag{11}$$
$$A \cap B \subset A. \tag{12}$$

If 0 denotes the vacuous set, then

$$A \cup 0 = A, \qquad A \cap 0 = 0. \tag{13}$$

If $A \subset B$, then

$$A \cup B = B, \qquad A \cap B = A. \tag{14}$$

2.14. Theorem. Let $\{E_n\}$, $n = 1, 2, 3, \ldots$, *be a sequence of countable sets, and put*

$$S = \bigcup_{n=1}^{\infty} E_n. \tag{15}$$

Then S is countable.

Proof: Let every set E_n be arranged in a sequence $\{x_{nk}\}$, $k = 1, 2, 3, \ldots$, and consider the infinite array

$$\begin{array}{ccccc} x_{11} & x_{12} & x_{13} & x_{14} & \cdots \\ x_{21} & x_{22} & x_{23} & x_{24} & \cdots \\ x_{31} & x_{32} & x_{33} & x_{34} & \cdots \\ x_{41} & x_{42} & x_{43} & x_{44} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{array} \tag{16}$$

in which the elements of E_n form the nth row. The array contains all elements of S. As indicated by the arrows, these elements can be arranged in a sequence

$$x_{11};\ x_{21},\ x_{12};\ x_{31},\ x_{22},\ x_{13};\ x_{41},\ x_{32},\ x_{23},\ x_{14};\ \ldots \tag{17}$$

If any two of the sets E_n have elements in common, these will appear more than once in (17). Hence there is a subset T of the set of all positive integers such that $S \sim T$, which shows that S is at most countable (Theorem 2.10). Since $E_1 \subset S$, and E_1 is infinite, S is infinite, and thus countable.

Corollary. *Suppose A is at most countable, and, for every $\alpha \in A$, B_α is at most countable. Put*

$$T = \bigcup_{\alpha \in A} B_\alpha.$$

Then T is at most countable.

For T is equivalent to a subset of (15).

2.15. Theorem. *Let A be a countable set, and let B_n be the set of all n-tuples $(a_1, \ldots, a_n)$, where $a_k \in A$ $(k = 1, \ldots, n)$, and the elements $a_1, \ldots, a_n$ need not be distinct. Then B_n is countable.*

Proof: That B_1 is countable is evident, since $B_1 = A$. Suppose B_{n-1} is countable ($n = 2,3,4, \ldots$). The elements of B_n are of the form

$$(b,a) \qquad (b \in B_{n-1},\ a \in A). \tag{18}$$

For every fixed b, the set of pairs (b,a) is equivalent to A, and hence countable. Thus B_n is the union of a countable set of countable sets. By Theorem 2.14, B_n is countable.

The theorem follows by induction.

Corollary. *The set of all rational numbers is countable.*

Proof: We apply Theorem 2.15, with $n = 2$, noting that every rational r is of the form b/a, where a and b are integers. The set of pairs (a,b), and therefore the set of fractions b/a, is countable.

In fact, even the set of all algebraic numbers is countable (see Exercise 6).

That not all infinite sets are, however, countable, is shown by the next theorem.

2.16. Theorem. *Let A be the set of all sequences whose elements are the digits* 0 *and* 1. *This set A is uncountable.*

The elements of A are sequences like $1, 0, 0, 1, 0, 1, 1, 1, \ldots$.

Proof: Let E be a countable subset of A, and let E consist of the sequences $s_1, s_2, s_3, \ldots$. We construct a sequence s as follows. If the nth digit in s_n is 1, we let the nth digit of s be 0, and vice versa. Then the sequence s differs from every member of E in at least one place; hence $s \notin E$. But clearly $s \in A$, so that E is a proper subset of A.

We have shown that every countable subset of A is a proper subset of A. It follows that A is uncountable (for otherwise A would be a proper subset of A, which is absurd).

The idea of the above proof was first used by Cantor, and is called Cantor's diagonal process; for, if the sequences $s_1, s_2, s_3, \ldots$ are placed in an array like (16), it is the elements on the diagonal which are involved in the construction of the new sequence.

The reader who is familiar with the binary representation of the real numbers (base 2 instead of 10) will notice that Theorem 2.16 implies that the set of all real numbers is uncountable. We shall give a second proof of this fact in Theorem 2.40.

Metric Spaces

2.17. Definition. A set X, whose elements we shall call points, is said to be a metric space if with any two points p and q of X there is associated a real number $d(p,q)$, called the distance from p to q, such that

(*a*) $d(p,q) > 0$ if $p \neq q$; $d(p,p) = 0$;

(*b*) $d(p,q) = d(q,p)$;

(*c*) $d(p,q) \leq d(p,r) + d(r,q)$, for any $r \in X$.