CHAPTER 1

INTRODUCTION

1. Sequences and limits

A *sequence* $\{x_n\}$ is a collection of objects occurring in order; thus there is a first member x_1, a second member x_2, and so on indefinitely. For every positive integer k, there is a corresponding kth member of the sequence. The members of such a sequence need not be all different. We can have a sequence all of whose members are the same; such a sequence is called a constant sequence.

If $\{k_n\}$ is a strictly increasing sequence of positive integers, the sequence $\{x_{k_n}\}$ is called a *subsequence* of $\{x_n\}$. The definition implies that $\{x_n\}$ is a subsequence of itself.

A sequence $\{x_n\}$ of real numbers is said to converge to the limit x if, for every positive value of ϵ, all but a finite number of members of the sequence lie between $x-\epsilon$ and $x+\epsilon$. If a sequence $\{x_n\}$ of real numbers converges, every subsequence converges to the same limit. A sequence of real numbers which converges to zero is called a *null-sequence*. Thus if $\{x_n\}$ converges to x, the sequence $\{x_n - x\}$ is a null-sequence.

It is often convenient to represent real numbers by points on a line, and to speak of the point of abscissa x simply as the point x. The distance between the points x and y is $|x-y|$. To say that the sequence of real numbers $\{x_n\}$ converges to x is thus the same thing as saying that the sequence of points $\{x_n\}$ converges to the point x, or that the distance between the point x_n and the point x tends to zero as $n \to \infty$.

In the same way, we frequently speak of the complex number z as the point z in the complex plane. That the sequence $\{z_n\}$ of complex numbers converges to z means that the distance $|z_n - z|$ between the point z_n and the point z in the complex plane tends to zero as $n \to \infty$. This geometrical language is very convenient, but in fact very little geometry is used. All we need in analysis is that the distance between two points satisfies the triangle inequality $|z_1 - z_2| \leqslant |z_1 - z_3| + |z_3 - z_2|$.

As the complex plane is merely a geometrical picture, we are not compelled to use the Euclidean formula $|z_1 - z_2|$ for the distance between two points. In the extended complex plane, that is, the ordinary complex plane with an added ideal point at infinity, it is sometimes more convenient to use as the 'distance' between the points z_1 and z_2 the expression

$$\frac{2|z_1 - z_2|}{\sqrt{\{(1+|z_1|^2)(1+|z_2|^2)\}}}.$$

This is known as the chordal distance since it is the length of the chord joining the points corresponding to z_1 and z_2 on the Riemann sphere whose stereographic projection is the extended complex plane. This 'distance' satisfies the triangle inequality; and when we are dealing with convergence to a finite limit, it does not matter which definition of 'distance' we use.

There are, however, important differences. The complex plane is unbounded, that is, we can find pairs of points whose distances apart are as large as we like. But the extended complex plane with the chordal definition of 'distance' is bounded, since no pair of points is at a 'distance' apart which exceeds 2. The 'distance' from z to the point at infinity is $2/\sqrt{(1+|z|^2)}$. To say that the sequence $\{z_n\}$ tends to the limit ∞ is equivalent to saying that, for every positive number ϵ (< 2), all but a finite number of members of the sequence are at a 'distance' less than ϵ from the point at infinity; it is readily seen that this condition reduces to

$$|z_n| > \left\{\frac{4}{\epsilon^2} - 1\right\}^{\frac{1}{2}}$$

for all but a finite number of values of n, as it should since we can choose ϵ (< 2) so that $(4\epsilon^{-2} - 1)^{\frac{1}{2}}$ is as large as we please.

A sequence might be a sequence $\{x_n(t)\}$ of real functions, each continuous on an interval $a \leqslant t \leqslant b$. If this sequence converged for each value of t in the interval, its limit $x(t)$ would be a function defined on the same interval; $\{x_n(t)\}$ is then said to be pointwise convergent. The sequence is said to converge uniformly to $x(t)$ on the interval if, for every positive value of ϵ, there exists an integer n_0, depending on ϵ but not on t, such that, whenever $n \geqslant n_0$, the inequality

$$|x_n(t) - x(t)| < \epsilon$$

holds at all points of the interval; if this condition is satisfied, the limit-function $x(t)$ is itself continuous.

There is another definition of uniform convergence equivalent to this which depends on the idea of supremum (least upper bound). Suppose that, for every positive integer n, the function $|x_n(t) - x(t)|$ has supremum M_n. This means that

$$|x_n(t) - x(t)| \leqslant M_n \quad (a \leqslant t \leqslant b),$$

and that, for every positive value of ϵ, there exists a number t_0 of the interval such that

$$|x_n(t_0) - x(t_0)| > M_n - \epsilon.$$

The sequence $\{x_n(t)\}$ converges uniformly to $x(t)$ if and only if M_n tends to zero as $n \to \infty$.

This alternative definition of uniform convergence enables us to express the idea in a geometrical form. We call each function $x(t)$, continuous on $a \leqslant t \leqslant b$, a 'point'; and we define the 'distance' between two 'points' $x(t)$ and $y(t)$ as the supremum of $|x(t) - y(t)|$, a definition which satisfies the triangle inequality. To say that the sequence of continuous functions $\{x_n(t)\}$ converges uniformly to the continuous function $x(t)$ is equivalent to saying that the 'distance' from the 'point' $x_n(t)$ to the 'point' $x(t)$ tends to zero as $n \to \infty$.

We might have defined the 'distance' between the 'points' $x(t)$ and $y(t)$ as

$$\left\{\int_a^b |x(t) - y(t)|^2\, dt\right\}^{\frac{1}{2}}.$$

When the sequence of 'points' $\{x_n(t)\}$ converges to $x(t)$ with this definition of 'distance', the sequence of functions $\{x_n(t)\}$ is said to *converge in mean of order* 2 to $x(t)$. Uniform convergence evidently implies convergence in mean of order 2, but not conversely. For example, $\{nt/(1+n^2t^2)\}$ converges in mean of order 2 on $0 \leqslant t \leqslant 1$ to zero, but is not uniformly convergent.

The study of the properties of sets of 'points' in a 'space' whose only geometrical property is the existence of a 'distance' between each pair of 'points' is called *metric space topology*. It had its beginnings in the theory of linear sets of points, which arose in the 19th century from a discussion of the theory of

Fourier series and of the functions which can be represented by such series. An account of this will be found in the first volume of E. W. Hobson's *Theory of Functions of a Real Variable and the Theory of Fourier's Series* (Cambridge, 1921).

We could develop metric space topology as a piece of abstract mathematics, and never mention its applications. Here we shall consider the subject in relation to its applications and show how the ideas unify branches of analysis which may seem disconnected.

In the rest of this chapter, we introduce ideas which may be well-known to the reader but will be needed in the sequel, before proceeding to the definition of a metric space. A knowledge of the elements of classical analysis is assumed.

2. Sets

In the previous section, we used the word 'set'. It is unnecessary here to discuss the logical difficulties involved in the general notion of a set. By a set, we always mean a well-defined collection of distinct elements, called the members of the set. Sometimes we shall use the word 'class' or 'family' instead of 'set'.

A finite set has only a finite number of members. An infinite set has an infinite number of members. An infinite set of real numbers is not necessarily unbounded; for example, the set of all rational numbers between ± 1 is infinite but is bounded. A set is said to be *empty* or *void* if it has no members. For example, the set of all positive integers less than 10 is finite; the set of all positive integers is infinite; the set of all positive integers less than -2 is empty.

An infinite set is said to be *countable* or *denumerable* if its members can be arranged as a sequence. The set of all rational numbers x such that $0 < x \leqslant 1$ is countable. Consider first the rational numbers p/q, between 0 and 1, each expressed in its lowest terms; for each given q, arrange the numbers p/q in order of increasing p; then arrange the groups so obtained in order of increasing q. We get the sequence

$$\tfrac{1}{1}, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{4}, \tfrac{3}{4}, \tfrac{1}{5}, \tfrac{2}{5}, \tfrac{3}{5}, \tfrac{4}{5}, \tfrac{1}{6}, \tfrac{5}{6}, \tfrac{1}{7} \ldots$$

which proves the result. But it will be proved later that the set of all real numbers between 0 and 1 is not countable: the real numbers between 0 and 1 cannot be arranged as a sequence.

We said that a set is countable if its members can be arranged as a sequence. But a sequence is not a set, not even if all the members of the sequence are distinct. The reason for this is that a set is a collection of distinct entities, considered merely as a collection; but in a sequence the order is important. For example, from the set of positive integers, we can construct many different sequences; 1, 2, 3, 4, 5 ... and 1, 3, 2, 5, 7, 4, 9, 11, 6, ... are two different sequences containing all the positive integers.

Sets are usually denoted by capital letters, members of a set by small letters. The notation $a \in A$ means that a is a member of the set A; $a \notin A$ that a is not a member of A. The empty set—there is only one empty set—is denoted by $\varnothing$.

A finite set can be defined by writing down explicitly the elements which are its members. Thus $\{1, 3, 5, 7\}$ is a set with four members, the odd natural numbers less than 8. Again $\{0\}$ is the set whose only member is the number zero; it is not the empty set —it has one member.

Certain capital letters are reserved to denote particular sets. Some of these are

N: the set of natural numbers 0, 1, 2, 3, ...
R: the set of real numbers.
R^2: the set of all ordered pairs (x, y) of real numbers.
R^3: the set of all ordered triples (x, y, z) of real numbers.
$C[a, b]$: the set of all functions $x(t)$ continuous on $a \leqslant t \leqslant b$.

Sometimes it is convenient to define a set by stating a property which is possessed by all the members of the set and by no element which does not belong to the set. Thus the set $\{2, 3\}$ might be defined as the set of all real numbers x which satisfy the equation $x^2 - 5x + 6 = 0$; we should then denote the set by $\{x \in R: x^2 - 5x + 6 = 0\}$. Again the set $\{1, 3, 5, 7\}$ could be defined as $\{x \in N: x \equiv 1 \pmod 2,\ x < 8\}$. The set $\{x \in N: x \equiv 0 \pmod 2,\ x \equiv 1 \pmod 2\}$ is the empty set, since there is no natural number which is both even and odd.

3. Sets and subsets

The sets A and B are said to be equal, $A = B$, if every member of A belongs to B and every member of B belongs to A.

The set A is said to be contained in B if every member of A belongs to B. We then write $A \subseteq B$. In this case, we may also say that B contains A, and write $B \supseteq A$; or again that A is a subset of B. If $A = B$, then $A \subseteq B$ and $B \supseteq A$; and conversely.

The set A is said to be properly contained in B if every member of A belongs to B and there is at least one member of B which does not belong to A. We then write $A \subset B$ or $B \supset A$, and say that A is a proper subset of B.

Some authors write $A \subset B$ whenever A is a subset of B, not necessarily a proper subset of B, and do not use the notation $\subseteq, \supseteq$.

Evidently,

$$\text{if } A \subseteq B \text{ and } B \subseteq C, \text{ then } A \subseteq C;$$
$$\text{if } A \subseteq B \text{ and } B \subset C, \text{ then } A \subset C;$$
$$\text{if } A \subset B \text{ and } B \subseteq C, \text{ then } A \subset C;$$
$$\text{if } A \subset B \text{ and } B \subset C, \text{ then } A \subset C.$$

The empty set $\varnothing$ is a subset of every set. For if not, there would exist a set A of which $\varnothing$ is not a subset; this would mean that there is a member of $\varnothing$ which does not belong to A. This is impossible, since $\varnothing$ has no members.

4. The complement of a set

Let us now consider sets which are all subsets of some fixed set E. The complement of a subset A with respect to E is the set of members of E which do not belong to A. We shall denote it by A'. The following results are evident from the definition:

(i) $E' = \varnothing$;

(ii) $\varnothing' = E$;

(iii) $(A')' = A$;

(iv) $A \subseteq B$ if and only if $B' \subseteq A'$.

This notation is satisfactory so long as we are considering only complements with respect to one fixed set E. When complements with respect to several sets occur, other notations are sometimes used, such as $\complement_E A$, or $E - A$.

5. Finite unions and intersections

The union of a family of subsets of the fixed set E is defined to be the set of members of E, each of which belongs to at least one subset of the family. The intersection of a family of subsets is defined to be the set of members of E, each of which belongs to all the subsets of the family. In either case, the family may consist of a finite or of a infinite number of subsets of E; and if it consists of an infinite number of subsets, the family is not necessarily countable.

The *union* $A \cup B$ of two subsets A and B of the fixed set E is the subset of E which consists of those members of E which belong to A or to B or to both.

The *intersection* $A \cap B$ is the subset of E which consists of those members of E which belong to A and to B. If $A \cap B = \varnothing$, A and B have no common member and are said to be *disjoint*.

The intersection $A \cap B'$ of A and the complement of B with respect to E evidently consists of the members of A which do not belong to B. It is called the *relative complement* of B with respect to A.

If A, B, C are three subsets of E, then $(A \cup B) \cup C = A \cup (B \cup C)$.

For if $x \in (A \cup B) \cup C$, then $x \in A \cup B$, or $x \in C$, or x belongs to both. If $x \in C$, $x \in B \cup C$ and therefore $x \in A \cup (B \cup C)$. If $x \in A \cup B$, then x belongs to A or to B or to both; if $x \in A$, $x \in A \cup (B \cup C)$; if $x \in B$, then $x \in B \cup C$ and therefore $x \in A \cup (B \cup C)$. Thus if

$$x \in (A \cup B) \cup C, \quad \text{then} \quad x \in A \cup (B \cup C),$$

and so

$$(A \cup B) \cup C \subseteq A \cup (B \cup C).$$

Similarly $A \cup (B \cup C) \subseteq (A \cup B) \cup C$. Hence the result. We may therefore leave out the brackets and write $A \cup B \cup C$ for this triple union. And, evidently, the triple union does not depend on the order of the sets A, B, C.

If n is any positive integer and $A_1, A_2, \ldots, A_n$ are subsets of E, the set

$$A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$$

consists of the members of E which belong to at least one of the subsets A_i. It is denoted by

$$\bigcup_{i=1}^{n} A_i \quad \text{or by} \quad \bigcup_{i} A_i.$$

If A, B, C are three subsets of E, then $(A \cap B) \cap C = A \cap (B \cap C)$. The proof is similar to that given above. Again we may leave out the brackets, and write simply $A \cap B \cap C$. The intersection

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap A_3 \cap \ldots A_n$$

consists of those members of E which belong to all the subsets $A_1, A_2, \ldots, A_n$.

If A and B are any two subsets of E, then $(A \cup B)' = A' \cap B'$ *and* $(A \cap B)' = A' \cup B'$.

If $x \in (A \cup B)'$, $x \notin A \cup B$. Hence $x \notin A$ and therefore $x \in A'$. Similarly $x \in B'$, and so $x \in A' \cap B'$. Thus $(A \cup B)' \subseteq A' \cap B'$. Again, if $x \in A' \cap B'$, $x \in A'$ and so $x \notin A$; similarly $x \notin B$. Hence $x \notin A \cup B$, that is, $x \in (A \cup B)'$. Thus $A' \cap B' \subseteq (A \cup B)'$. The result follows.

The second result can be proved in the same way. Alternatively, it can be proved by taking complements. Remembering that $A'' = A$, we have $A \cup B = (A \cup B)'' = (A' \cap B')'$. Writing $A' = C$, $B' = D$, we obtain $A = A'' = C'$, $B = B'' = D'$; and so $C' \cup D' = (C \cap D)'$, as required.

6. Infinite unions and intersections

Let S be any set, finite or infinite; it might be the positive integers between 0 and 10, it might be all the positive integers, it might even be all the real numbers between 0 and 1. Suppose that, corresponding to each member of α of S, there exists a subset of E which we denote by A_α. We then have a family of subsets of E with index set S. We denote the union and intersection of the family by

$$\bigcup_{\alpha \in S} A_\alpha, \quad \bigcap_{\alpha \in S} A_\alpha.$$

Evidently

$$\Big(\bigcup_{\alpha \in S} A_\alpha\Big)' = \bigcap_{\alpha \in S} A'_\alpha, \quad \Big(\bigcap_{\alpha \in S} A_\alpha\Big)' = \bigcup_{\alpha \in S} A'_\alpha.$$

The union of a countable family of countable sets is countable.

A countable family of sets can be arranged as a sequence $A_1, A_2, A_3, \ldots$. Let $A_m = \{a_{m,1}, a_{m,2}, a_{m,3}, \ldots\}$. Then $\bigcup_{m=1}^{\infty} A_m$ consists of all elements $a_{m,n}$, and these can be arranged as a sequence

$$a_{1,1}, a_{1,2}, a_{2,1}, a_{1,3}, a_{2,2}, a_{3,1}, a_{1,4}, \ldots.$$

If the sets are not disjoint, there will be repetitions in this sequence. But the elements of $\bigcup_{m=1}^{\infty} A_m$ form a subsequence of this sequence when repetitions are omitted. Hence $\bigcup_{m=1}^{\infty} A_m$ is countable.

It follows from this that the set of all rational numbers is countable. We have seen that the set A_0 consisting of all rational numbers x, such that $0 < x \leqslant 1$ is countable. It follows that the set A_1, consisting of all rational numbers x such that $1 < x \leqslant 2$ is countable since they are in one-one correspondence with the members of A_0. Similarly for A_n, the set of all rational numbers x such that $n < x \leqslant n+1$ for any positive or negative integer n. The set of all rational numbers is

$$A_0 \cup A_1 \cup A_{-1} \cup A_2 \cup A_{-2} \cup \ldots;$$

hence the result.

7. The algebra of sets

Let A, B, C be subsets of a given set E. Then the operations of union, intersection and complementation have the following properties:

(*a*) $A \cup A' = E$, $\quad A \cap A' = \varnothing$.

(*b*) $A \cup A = A$, $\quad A \cap A = A$.

(*c*) $A \cup \varnothing = A$, $\quad A \cap E = A$.

(*d*) $A \cap \varnothing = \varnothing$, $\quad A \cup E = E$.

(*e*) $A \cup B = B \cup A$, $\quad A \cap B = B \cap A$.

(*f*) $(A \cup B) \cup C = A \cup (B \cup C)$, $\quad (A \cap B) \cap C = A \cap (B \cap C)$.

(*g*) $A \subseteq A \cup B$, $\quad A \supseteq A \cap B$.

(*h*) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,

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

These properties have been arranged in pairs; each relation in a pair can be deduced from the other by using the duality property of the previous section, viz. $(A \cup B)' = A' \cap B'$ and its equivalent $(A \cap B)' = A' \cup B'$. Thus it is necessary only to prove half these formulae; in fact, many of them are obvious from the definitions of union and intersection.

In (*g*), it is evident that $A \subseteq A \cup B$. Hence

$$A' \supseteq (A \cup B)' = A' \cap B'.$$

Replace A', B' by C, D. Then $C \supseteq C \cap D$.

To prove the first formula in (*h*), note that, if $x \in (A \cup B) \cap C$, then x belongs to $A \cup B$ and to C. But if $x \in A \cup B$, x belongs to A or to B or to both; hence x belongs to $A \cap C$ or to $B \cap C$ or to both. Thus $x \in (A \cap C) \cup (B \cap C)$, and so

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

Again, if $x \in (A \cap C) \cup (B \cap C)$, x belongs to $A \cap C$ or to $B \cap C$ or to both. If $x \in A \cap C$, x belongs to A and to C; therefore $x \in A \cup B$ and $x \in C$ and so $x \in (A \cup B) \cap C$; and similarly if $x \in B \cap C$. Hence $(A \cap C) \cup (B \cap C) \subseteq (A \cup B) \cap C$, which completes the proof.

To obtain the second formula in (*h*), we take the complement of the first formula, which gives

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

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

or, replacing A', B', C' by their complements,

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

8. Equivalence relations

Two natural numbers a and b may be connected by one of many relations; for example, $a < b$, $a \equiv b \pmod 3$, a is a divisor of b, a is a multiple of b, and so on. All these are examples of what are called *binary relations*. Such relations occur in many other branches of mathematics. In plane geometry, the property that a point a is conjugate to a point b with respect to a fixed