Introduction

The concept of *distance* plays a crucial role in everyday life. We frequently need to know how far it is from one place to another, and important decisions are influenced by such considerations. Hence it is often necessary to have a notion of distance built into the mathematical theories which we use to model the various things that interest us.

The Euclidean plane provides the most familiar example of distance in a mathematical context. Using a Cartesian coordinate system enables us to identify points in the plane with elements of the set $\mathbb{R}^2 = \{\, (x,y) \mid x,\, y \in \mathbb{R} \,\}$ (where $\mathbb{R}$ is the set of all real numbers). By Pythagoras' Theorem the distance $d(a,b)$ between the points $a = (x_0, y_0)$ and $b = (x_1, y_1)$ in $\mathbb{R}^2$ is given by the formula

$$d(a,b) = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}.$$

The function d defined by this formula is known as the *Euclidean metric on* $\mathbb{R}^2$. We can use this Euclidean metric to define the concept of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}^2$:

> A function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is continuous at the point $a \in \mathbb{R}^2$ if for every real number $\varepsilon > 0$ there exists a real number $\delta > 0$ such that the following condition holds: for all $x \in \mathbb{R}^2$, if $d(x,a) < \delta$ then $d(f(x), f(a)) < \varepsilon$.

This definition will still make sense if $\mathbb{R}^2$ is replaced by any other set X, provided that X has a distance function d defined on it; thus sets equipped with distance functions provide a natural context in which to study continuity. To qualify as a distance function, d must satisfy the following three properties:

(1) for all $a,\, b \in X$ the distance $d(a,b)$ is a nonnegative real number, and $d(a,b) = 0$ if and only if $a = b$;
(2) $d(a,b) = d(b,a)$, for all $a,\, b \in X$;
(3) $d(a,b) + d(a,c) \geq d(b,c)$, for all $a,\, b,\, c \in X$.

A distance function is alternatively known as a *metric*, and a *metric space* is a pair (X,d) comprising a set X and a metric d defined on X.

Some background material

A *set* is a mathematical object made up of other mathematical objects, known as the *elements* or *members* of the set. The notation $x \in A$ says that x is an element of the set A. If A and B are sets then $A = B$ if and only if every element of A is an element of B and every element of B is an element of A. In other words, a set is completely determined by its elements. A set B is a *subset* of a set A if every element of B is an element of A. We shall write $B \subseteq A$ to indicate that B is a subset of A. We call B a *proper* subset of A if $B \subseteq A$ and $B \neq A$; we indicate this by writing $B \subset A$.† Sets which have only a finite number of elements are commonly specified by enclosing a list of the elements within a left brace ({) and a right brace (}); for example, $\{2, 5, 9\}$ is the set whose elements are the numbers 2, 5 and 9. But sets are usually specified by use of the

† This is a departure from the traditional—and most common—notation, which uses $B \subset A$ where we use $B \subseteq A$.

notation "$\{\ldots \mid \ldots\}$," which should be read as "the set of all ... such that ... ". For example, $\{ x \in \mathbb{R} \mid 0 < x < 4 \}$ is the set whose members are those real numbers x which satisfy the condition $0 < x < 4$.

If A and B are sets, their *intersection* $A \cap B$ and *union* $A \cup B$ are defined by

$$A \cap B = \{ x \mid x \in A \text{ and } x \in B \},$$
$$A \cup B = \{ x \mid x \in A \text{ or } x \in B \}.$$

(Note that, as is usual in mathematics, this "or" is the inclusive variety: p or q is true if p and q are both true, or if p is true and q false, or if p is false and q is true; it is false if p and q are both false.) We also define

$$B \setminus A = \{ x \in B \mid x \notin A \}$$

and

$$A \times B = \{ (a, b) \mid a \in A, \quad b \in B \}.$$

The set $A \times B$ is called the *Cartesian product* of A and B.

There is a unique set which has no elements; it is denoted by $\emptyset$. Various sets of numbers are commonly designated by "blackboard bold" letters, as follows:

$\mathbb{R}$ is the set of all real numbers,
$\mathbb{C}$ is the set of all complex numbers,
$\mathbb{Q}$ is the set of all rational numbers,
$\mathbb{Z}$ is the set of all integers,
$\mathbb{N}$ is the set of all nonnegative integers.†

For a, $b \in \mathbb{R}$ we often use the following notation:

$$[a, b] = \{ x \in \mathbb{R} \mid a \le x \le b \},$$
$$(a, b] = \{ x \in \mathbb{R} \mid a < x \le b \},$$
$$[a, b) = \{ x \in \mathbb{R} \mid a \le x < b \},$$
$$(a, b) = \{ x \in \mathbb{R} \mid a < x < b \}.$$

Note that, for example, (a, b) is empty if $a \ge b$. (It is unfortunate that the notation (a, b) is used both for ordered pairs (elements of $A \times B$) and intervals of real numbers; it is often necessary to deduce from the context which convention is being used.)

Let X and Y be sets and $f\colon X \to Y$ a function. For each $A \subseteq X$ we define the *image* of A under f to be the set

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

(a subset of Y). For each $B \subseteq Y$ we define the *preimage* or *inverse image* of B in X to be the set

$$f^{-1}(B) = \{ x \in X \mid f(x) \in B \}.$$

The following result is trivial.

Lemma. *Let $f\colon X \to Y$ be a function, and let $A \subseteq X$ and $B \subseteq Y$. Then $f(A) \subseteq B$ if and only if $A \subseteq f^{-1}(B)$.*

There are several similar properties of images and preimages discussed on p. 3 of Dr Choo's notes. Please read this.

We shall recall more background material as we proceed.

---

† Many authors use $\mathbb{N}$ for the set of all positive integers. (That is, they exclude zero from $\mathbb{N}$.) We shall use $\mathbb{Z}^+$ for the positive integers.

Some definitions and preliminary results

Definition. Let X be a set. A *metric* on X is a function $d: X \times X \to \mathbb{R}$ such that for all a, b, $c \in X$,

(M1) $d(a,b) = d(b,a) \geq 0$,

(M2) $d(a,b) = 0$ if and only if $a = b$, and

(M3) $d(a,b) + d(a,c) \geq d(b,c)$.

A *metric space* is a pair (X,d), where X is a set and d a metric on X.

It is quite possible for a set X to have two different metrics d and d' defined on it, and then the metric space (X,d) is different from the metric space (X,d'). However, we are usually only concerned with one metric at a time, and so we often (loosely) talk of "the metric space X", since the metric is clear from the context.

Definition. Let (X,d) be a metric space, let $a \in X$ and let $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$. The *open ball* with centre at a and radius ε is the set

$$B_d(a,\varepsilon) = \{\, x \in X \mid d(x,a) < \varepsilon \,\}.$$

When the metric is clear from the context, we often write $B(a,\varepsilon)$ instead of $B_d(a,\varepsilon)$.

Definition. Let (X,d) be a metric space and let $S \subseteq X$. An element $a \in S$ is called an *interior point* of S if there is a positive real number ε with $B(a,\varepsilon) \subseteq S$. The *interior* of S is the set $\operatorname{Int}(S)$ consisting of all interior points of S, and S is said to be an *open set* if $S = \operatorname{Int}(S)$ (that is, if all its points are interior points).

Definition. Let (X,d) and (Y,d') be metric spaces. A function $f: X \to Y$ is said to be *continuous* at the point $a \in X$ if for all $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x \in X$, if $d(x,a) < \delta$ then $d'(f(x), f(a)) < \varepsilon$. If f is continuous at all points $a \in X$ then it is said to be *continuous*.

As we shall soon see, it is possible to utilize the concept of "open set", defined above, to give a briefer characterization of continuity. First, in justification of the terminology we have introduced, let us prove that open balls are open sets.

Lemma. *Let (X,d) be a metric space and $a \in X$. For every positive real number t, the set $B(a,t)$ is an open subset of X.*

Proof. We must show that every point $b \in B(a,t)$ is an interior point of $B(a,t)$. That is, we must show that if $b \in B(a,t)$ then there is a $\delta > 0$ such that $B(b,\delta) \subseteq B(a,t)$.

Let $b \in B(a,t)$. Then $d(b,a) < t$, and so δ defined by $\delta = t - d(b,a)$ is positive. Now if $x \in B(b,\delta)$ is arbitrary, then $d(b,x) < \delta$, and so

$$d(x,a) \leq d(b,x) + d(b,a) < \delta + d(b,a) = t \tag{1}$$

by the definition of δ. Since (1) implies that $x \in B(a,t)$, we have shown that every element of $B(b,\delta)$ is an element of $B(a,t)$, as required. $\square$

The key point in the above proof is, of course, that for all points in $B(a,t)$ the distance to a is strictly less than t; points whose distance from a is exactly t are not in $B(a,t)$. But if b has the property that its distance from a is strictly less than t then all points close enough to b will also have this same property. Points on the boundary of $B(a,t)$—points whose distance from a is t—are not in $B(a,t)$; all points which are in $B(a,t)$ are in its interior, because they are surrounded, so to speak, by other elements of $B(a,t)$. Open sets can be thought of as those sets for which points on the boundary are not in the set. (We shall give a precise definition of the term "boundary point" later on.)