# Elementary Point Set Topology

## 1.2 Metric Spaces

**Definition** A set $S$ is a *metric space* if there is a real-valued function $d : S \times S \to \mathbb{R}$, called a metric (or distance) function, such that for all $x,\, y,\, z \in S$:

1. $d(x, y) \geq 0$ and equality holds if and only if $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, y) + d(y, z) \geq d(x, z)$. (i.e. the triangle inequality holds.)

Given a metric $d$ on $S$, any $y \in S$ and any $\delta > 0$, the set $B(y,\, \delta)$ defined by

$$B(y,\, \delta) = \{x \in S \mid d(x, y) < \delta\}$$

is called the *ball* with center $y$ and radius $\delta$.

**Example** $\mathbb{R}$, $\mathbb{C}$ and the $n$-dimensional euclidean space $\mathbb{R}^n$ are metric spaces with $d(x,\, y) = |x - y|$, the euclidean distance between $x$ and $y$.

The extended complex plane $\widehat{\mathbb{C}}$ is a metric space with the (bounded) metric defined by

$$d(z,\, z') = \frac{2|z - z'|}{\sqrt{(1 + |z|^2)\,(1 + |z'|^2)}} \quad \text{for } z,\, z' \in \widehat{\mathbb{C}}$$

**Definition** 1. A set $N \subset S$ is called a *neighborhood of $y \in S$* if it contains a ball $B(y,\, \delta)$.

− In other words, a neighborhood of $y$ is a set which contains all points sufficiently near to $y$.

**Definition** 2. A set is *open* if it is a neighborhood of each of its elements, i.e. $N \subset S$ is open if for each $y \in N$, there exists a ball $B(y,\, \delta) \subset N$.

− *Every ball is an open set.*

Indeed, if $z \in B(y,\, \delta)$, then $\delta' = \delta - d(y,\, z) > 0$. The triangle inequality shows that $B(z,\, \delta') \subset B(y,\, \delta)$, for $d(x,\, z) < \delta'$ gives $d(x,\, y) \leq d(x,\, z) + d(z,\, y) < \delta' + d(z,\, y) = \delta$. Hence $B(y,\, \delta)$ is a neighborhood of $z$, and since $z$ was any point in $B(y,\, \delta)$ we conclude that $B(y,\, \delta)$ is an open set. For greater emphasis a ball is sometimes referred to as an *open ball*, to distinguish it from the *closed ball* formed by all $x \in S$ with $d(x,\, y) \leq \delta$.

In the complex plane $B(z_0,\, \delta)$ is an open disk with center $z_0$ and radius $\delta$; it consists of all complex numbers $z$ which satisfy the strict inequality $|z - z_0| < \delta$.

The complement of an open set is said to be *closed*. In any metric space the empty set and the whole space are at the same time open and closed, and there may be other sets with the same property.

The following properties of open and closed sets are fundamental:

− *The intersection of a finite number of open sets is open.*

− *The union of any collection of open sets is open.*

− *The union of a finite number of clo-sed sets is closed.*

− *The intersection of any collection of closed sets is closed.*

## More Related Definitions

There are many terms in common usage which are directly related to the idea of open sets. A complete list would be more confusing than helpful, and we shall limit ourselves to the following: interior, closure, boundary, exterior.

– The *interior* of a set $X$ is the largest open set contained in $X$. It exists, for it may be characterized as the union of all open sets contained in $X$. It can also be described as the set of all points of which $X$ is a neighborhood. We denote it by $\text{Int}\, X$.

– The *closure* of $X$ is the smallest closed set which contains $X$, or the intersection of all closed sets containing $X$. A point belongs to the closure of $X$ if and only if all its neighborhoods intersect $X$. The closure is usually denoted by $\bar{X}$ or $\text{Cl}\, X$.

– The *boundary* of $X$ is the closure minus the interior. A point belongs to the boundary if and only if all its neighborhoods intersect both $X$ and the *complement* of $X$, denoted by $\sim X$ or $X^c$. Notation: $\partial X$ or $\text{Bd}\, X$.

– The *exterior* of $X$ is the interior of $X^c$. It is also the complement of the closure. As such it can be denoted by $\bar{X}^c$ or $\sim \bar{X}$.

Observe that $\text{Int}\, X \subset X \subset \bar{X}$ and that $X$ is open if $\text{Int}\, X = X$, closed if $\bar{X} = X$. Also, $X \subset Y$ implies $\text{Int}\, X \subset \text{Int}\, Y$, $\bar{X} \subset \bar{Y}$. For added convenience we shall also introduce the notions of isolated point and accumulation point. We say that $x \in X$is an *isolated point* of $X$ if $x$ has a neighborhood whose intersection with $X$ reduces to the point $x$. An *accumulation point* is a point of $\bar{X}$ which is not an isolated point. It is clear that $x$ is an accumulation point of $X$ if and only if every neighborhood of $x$ contains infinitely many points from $X$.

### 1.3 Connectedness

**Definition** 3. A metric space $S$ is said to be *disconnected* if there exists a partition $S = A \cup B$ into disjoint union of nonempty open subsets $A$ and $B$.

A space $S$ is said to be *connected* if it cannot be represented as the disjoint union of two nonempty open sets.

**Remark** A subset $E \subset S$ is said to be connected if it is connected in the relative topology, i.e. a subset of a metric space is connected if it cannot be represented as the disjoint union of two nonempty relatively open sets.

If $E$ is open, a subset of $E$ is relatively open if and only if it is open. Similarly, if $E$ is closed, relatively closed means the same as closed. We can therefore state: *An open set is connected if it cannot be decomposed into two open sets, and a closed set is connected if it cannot be decomposed into two closed sets.* Again, none of the sets is allowed to be empty.

Trivial examples of connected sets are the empty set and any set that consists of a single point.

In the case of the real line it is possible to name all connected sets. The most important result is that the whole line is connected, and this is indeed one of the fundamental properties of the real-number system.

An *interval* is defined by an inequality of one of the four types: $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$, $[a, b) = \{x \in \mathbb{R} \mid a \le x < b\}$, $(a, b] = \{x \in \mathbb{R} \mid a < x \le b\}$, $[a, b] = \{x \in \mathbb{R} \mid a \le x \le b\}$. For $a = -\infty$ or $b = \infty$ this includes the semi-infinite intervals $(-\infty, b)$, $(a, \infty)$ and the whole line $(-\infty, \infty)$.

**Theorem** 1. The nonempty connected subsets of the real line are the intervals.

**Proof** Suppose that the real line $\mathbb{R}$ is represented as the union $s \in \mathbb{R} = A \cup B$ of two disjoint closed sets. If neither is empty we can find $a_1 \in A$ and $b_1 \in B$; we may assume that $a_1 < b_1$. We bisect the interval $(a_1, b_1)$ and note that one of the two halves has its left end point in $A$ and its right end point in $B$. We denote this interval by $(a_2, b_2)$ and continue the process indefinitely. In this way we obtain a sequence of nested intervals $(a_n, b_n)$ with $a_n \in A$ and $b_n \in B$. The sequences $\{a_n\}$ and $\{b_n\}$ have a common limit $c$. Since $A$ and $B$ are closed $c$ would have to be a common point of $A$ and $B$. This contradiction shows that either $A$ or $B$ is empty, and hence $\mathbb{R}$ is connected. $\square$

With minor modifications the same proof applies to any interval.

**Definition** Let $E$ be an arbitrary subset of $\mathbb{R}$ and call $\alpha$ a *lower bound* of $E$ $\alpha \leq x$ for all $x \in E$. Consider the set $A$ of all lower bounds of $E$. The largest number $a$ of $A$, if it exists, is called *the greatest lower bound* of $E$; it is commonly denoted as $\text{g.l.b.}\, x$ or $\inf x$ for $x \in E$. If $A$ is empty, we agree to set $a = -\infty$, and if $A$ is the whole line we set $a = \infty$. The *least upper bound*, denoted as $\text{l.u.b.}\, x$ or $\sup x$ for $x \in E$, is defined in a corresponding manner.

Returning to the proof, we assume that $E$ is a connected set with the greatest lower bound $a$ and the least upper bound $b$. All points of $E$ lie between $a$ and $b$, limits included. Suppose that a point $\xi$ from the open interval $(a, b)$ did not belong to $E$. Then the open sets defined by $x < \xi$ and $x > \xi$ cover $E$, and because $E$ is connected, one of them must fail to meet $E$. Suppose, for instance, that no point of $E$ lies to the left of $\xi$. Then $\xi$ would be a lower bound, in contradiction with the fact that $a$ is the greatest lower bound. The opposite assumption would lead to a similar contradiction, and we conclude that $\xi$ must belong to $E$. It follows that $E$ is an open, closed, or semiclosed interval with the end points $a$ and $b$; the cases $a = -\infty$ and $b = \infty$ are to be included.

In the course of the proof we have introduced the notions of greatest lower bound and least upper bound. If the set is closed and if the bounds are finite, they must belong to the set, in which case they are called the minimum and the maximum. In order to be sure that the bounds are finite we must know that the set is not empty and that there is some finite lower bound and some finite upper bound. In other words, the E¿et must lie in a finite interval; such a set is said to be bounded. We have proved:

**Theorem 2.** Any closed and bounded nonempty set of real numbers has a minimum and a maximum.

The structure of connected sets in the plane is not nearly so simple as in the case of the line, but the following characterization of open connected sets contains essentially all the information we shall need.

**Theorem 3.** A nonempty open set in the plane is connected if and only if any two of its points can be joined by a polygon which lies in the set.

**Proof** We prove first that the condition is necessary. Let $A$ be an open connected set, and choose a point $a \in A$. We denote by $A_1$ the subset of $A$ whose points can be joined to a by polygons in $A$, and by $A_2$ the subset whose points cannot be so joined. Let us prove that $A_1$ and $A_2$ are both open. First, if $a_1 \in A_1$ there exists a neighborhood $|z - a_1| < \varepsilon$ contained in $A$. All points in this neighborhood can be joined to $a_1$ by a line segment, and from there to $a$ by a polygon. Hence the whole neighborhood is contained in $A_1$ and $A_1$ is open. Secondly, if $a_2 \in A_2$, let $|z - a_1| < \varepsilon$ be a neighborhood contained in $A$. If a point in this neighborhood could be joined to $a$ by a polygon, then $a_2$ could be joined to this point by a line segment, and from there to $a$. This is contrary to the definition of $A_2$, and we conclude that $A_2$ is open. Since $A$ was connected either $A_1$ or $A_2$ must be empty. But $A_1$ contains the point $a$; hence $A_2$ is empty, and all points can be joined to $a$. Finally, any two points in $A$ can be joined by way of $a$, and we have proved that the condition is necessary.

For future use we remark that it is even possible to join any two points by a polygon whose sides are parallel to the coordinate axes. The proof is the same.

In order to prove the sufficiency we assume that $A$ has a representation $A = A_1 \cup A_2$ as the union of two disjoint open sets. Choose $a_1 \in A_1$, $a_2 \in A_2$ and suppose that these points can be joined by a polygon in $A$.

One of the sides of the polygon must then join a point in $A_1$ to a point in $A_2$, and for this reason it is sufficient to consider the case where $a_1$ and $a_2$ are joined by a line segment. This segment has a parametric representation $z = a_1 + t(a_2 - a_1)$ where $t$ runs through the interval $1 \leq t \leq 1$. The subsets of the interval $0 < t < 1$ which correspond to points in $A_1$ and $A_2$, respectively, are evidently open, disjoint, and nonvoid. This contradicts the connectedness of the interval, and we have proved that the condition of the theorem is sufficient. □

The theorem generalizes easily to $\mathbb{R}^n$ and $\mathbb{C}^n$.

**Definition 4.** A nonempty connected open set is called a *region*. The closure of a region is called a *closed region*.

A *component* of a set is a connected subset which is not contained in any larger connected subset.

**Theorem 4.** Every set has a unique decomposition into components.

**Proof** If $E$ is the given set, consider a point $a \in E$ and let $C(a)$ denote the union of all connected subsets of $E$ that contain $a$. Then $C(a)$ is sure to contain $a$, for the set consisting of the single point $a$ is connected. If we can show that $C(a)$ is connected, then it is a maximal connected set, in other words a component. It would follow, moreover, that any two components are either disjoint or identical, which is precisely what we want to prove. Indeed, if $c \in C(a) \cap C(b)$, then $C(a) \subset C(c)$ by the definition of $C(c)$ and the connectedness of $C(a)$. Hence $a \in C(c)$ and by the same reasoning $C(c) \subset C(a)$, so that in fact $C(a) = C(c)$. Similarly $C(b) = C(c)$, and consequently $C(a) = C(b)$. We call $C(a)$ the *component of $a$*.

Suppose that $C(a)$ were not connected. Then we could find relatively open sets $A, B \neq \emptyset$ such that $C(a) = A \cup B$, $A \cap B = \emptyset$. We may assume that $a \in A$ while $B$ contains a point $b$. Since $b \in C(a)$ there is a connected set $E_0 \subset E$ which contains $a$ and $b$. The representation $E_0 = (E_0 \cap A) \cup (E_0 \cap B)$ would be a decomposition into relatively open subsets, and since $a \in (E_0 \cap A)$, $b \in (E_0 \cap B)$, neither part would be empty. This is a contradiction, and we conclude that $C(a)$ is connected. □

**Theorem 5.** In $\mathbb{R}^n$ the components of any open set are open.

This is a consequence of the fact that the $\delta$-neighborhoods in $\mathbb{R}^n$ are connected. Consider $a \in C(a) \subset E$. If $E$ is open it contains $B(a, \delta)$ and because $B(a, \delta)$ is connected $B(a, \delta) \subset C(a)$. Hence $C(a)$ is open. A little more generally the assertion is true for any space $S$ which is *locally connected* if any neighborhood of a point $a \in S$ contains a connected neighborhood of $a$.

In the case of $\mathbb{R}^n$ we can conclude, furthermore, that the number of components is countable. To see this we observe that every open set must contain a point with rational coordinates. The set of points with rational coordinates is countable, and may thus be expressed as a sequence $\{p_k\}$. For each component $C(a)$, determine the smallest $k$ such that $p_k \in C(a)$. To different components correspond different $k$. We conclude that the components are in one-to-one correspondence with a subset of the natural numbers, and consequently the set of components is countable.

For instance, every open subset of $\mathbb{R}$ is a countable union of disjoint open intervals.

Again, it is possible to analyze the proof and thereby arrive at a more general result. We shall say that a set $E$ is *dense* in $S$ if $\bar{E} = S$, and we shall say that a metric space is *separable* if there exists a countable subset which is dense in $S$. We are led to the following result:

*In a locally connected separable space every open set is a countable union of disjoint regions.*

### 1.4 Compactness

The notions of *convergent* sequences and *Cauchy sequences* are obviously meaningful in any metric space. Indeed, we would say that $x_n \to x$ if $d(x_n, x) \to 0$, and we would say that $\{x_n\}$ is a Cauchy sequence if $d(x_n, x_m) \to 0$, as $n$ and $m$ tend to $\infty$. It is clear that every convergent sequence is a Cauchy sequence. For $\mathbb{R}$ and $\mathbb{C}$ we have proved the converse, namely that every Cauchy sequence is convergent, and it is not hard to see that this property carries over to any $\mathbb{R}^n$. In view of its importance the property deserves a special name.

**Definition** 5. A metric space is said to be *complete* if every Cauchy sequence is convergent.

A subset is complete if it is complete when regarded as a subspace. The reader will find no difficulty in proving that *a complete subset of a metric space is closed*, and that *a closed subset of a complete space is complete*.

We shall now introduce the stronger concept of *compactness*. It is stronger than completeness in the sense that every compact space or set is complete, but not conversely. As a matter of fact it will turn out that the compact subsets of $\mathbb{R}$ and $\mathbb{C}$ are the closed bounded sets. In view of this result it would be possible to dispense with the notion of compactness, at least for the purposes of this book, but this would be unwise, for it would mean shutting our eyes to the most striking property of bounded and closed sets of real or complex numbers. The outcome would be that we would have to repeat essentially the same proof in many different connections.

There are several equivalent characterizations of compactness, and it is a matter of taste which one to choose as definition. Whatever we do the uninitiated reader will feel somewhat bewildered, for he will not be able to discern the purpose of the definition. This is not surprising, for it took a whole generation of mathematicians to agree on the best approach. The consensus of present opinion is that it is best to focus the attention on the different ways in which a given set can be covered by open sets.

Let us say that a collection of open sets is an *open covering* of a set $X$ if $X$ is contained in the union of the open sets. A *subcovering* is a subcollection with the same property, and a *finite covering* is one that consists of a finite number of sets. The definition of compactness reads:

**Definition** 6. A set $X$ is compact if and only if every open covering of $X$ contains a finite subcovering.

In this context we are thinking of $X$ as a subset of a metric space $S$, and the covering is by open sets of $S$. But if $U$ is an open set in $S$, then $U \cap X$ is an open subset of $X$ (a relatively open set), and conversely every open subset of $X$ can be expressed in this form. For this reason it makes no difference whether we formulate the definition for a full space or for a subset.

The property in the definition is frequently referred to as the *Heine Borel property*. Its importance lies in the fact that many proofs become particularly simple when formulated in terms of open coverings.

We prove first that *every compact space is complete*. Suppose that $X$ is compact, and let $\{x_n\}$ be a Cauchy sequence in $X$. If $y$ is not the limit of $\{x_n\}$ there exists an $\varepsilon > 0$ such that $d(x_n, y) > 2\varepsilon$ for infinitely many $n$. Determine $n_0$ such that $d(x_m, x_n) < \varepsilon$ for $m, n \geq n_0$. We choose a fixed $n \geq n_0$ for which $d(x_n, y) > 2\varepsilon$. Then $d(x_m, y) \geq d(x_n, y) - d(x_m, x_n) > \varepsilon$ for all $m \geq n_0$. It follows that the $\varepsilon$-neighborhood $B(y, \varepsilon)$ contains only finitely many $x_n$ (better: contains $x_n$ only for finitely many $n$).

Consider now the collection of all open sets $U$ which contain only finitely many $x_n$. If $\{x_n\}$ is not convergent, it follows by the preceding reasoning that this collection is an open covering of $X$. Therefore it must contain a finite subcovering, formed by $U_1, \ldots, U_N$. But that is clearly impossible, for since each $U_i$ contains only finitely many $x_n$ it would follow that the given sequence is finite.

Secondly, *a compact set is necessarily bounded* (a metric space is *bounded* if all distances lie under a finite bound). To see this, choose a point $x_0$ and consider all balls $B(x_0, r)$. They form an open covering of $X$, and if $X$ is compact, it contains a finite subcovering; in other words, $X \subset B(x_0, r_1) \cup \cdots \cup B(x_0, r_m)$, which means the same as $X \subset B(x_0, r)$ with $r = \max(r_1, \ldots, r_m)$. For any $x, y \in X$ it follows that $d(x, y) < d(x, x_0) + d(y, x_0) < 2r$, and we have proved that $X$ is bounded.

But boundedness is not all we can prove. It is convenient to define a stronger property called total boundedness:

**Definition 7.** A set $X$ is *totally bounded* if, for every $\varepsilon > 0$, $X$ can be covered by finitely many balls of radius $\varepsilon$.

This is certainly true of any compact set. For the collection of all balls of radius $\varepsilon$ is an open covering, and the compactness implies that we can select finitely many that cover $X$. We observe that a totally bounded set is necessarily bounded, for if $X \subset B(x_1, \varepsilon) \cup \cdots \cup B(x_m, \varepsilon)$, then any two points of $X$ have a distance $2\varepsilon + \max d(x_i, x_j)$. (The preceding proof that any compact set is bounded becomes redundant.)

We have already proved one part of the following theorem:

**Theorem 6.** A set is compact if and only if it is complete and totally bounded.

To prove the other part, assume that the metric space $S$ is complete and totally bounded. Suppose that there exists an open covering which does not contain any finite subcovering. Write $\varepsilon_n = 2^{-n}$. We know that S can be covered by finitely many $B(x, \varepsilon_1)$. If each had a finite subcovering, the same would be true of $S$; hence there exists a $B(x_1, \varepsilon_1)$ which does not admit a finite subcovering. Because $B(x_1, \varepsilon_1)$ is itself totally bounded we can find an $x_2 \in B(x_1, \varepsilon_1)$ such that $B(x_2, \varepsilon_2)$ has no finite subcovering. It is clear how to continue the construction: we obtain a sequence $x_n$ with the property that $B(x_n, \varepsilon_n)$ has no finite subcovering and $x_{n+1} \in B(x_n, \varepsilon_n)$. The second property implies $d(x_n, x_{n+1}) < \varepsilon_n$ and hence $d(x_n, x_{n+p}) < \varepsilon_n + \varepsilon_{n+1} + \cdots + \varepsilon_{n+p-1} < 2^{-n+1}$. It follows that $x_n$ is a Cauchy sequence. It converges to a limit $y$, and this $y$ belongs to one of the open sets $U$ in the given covering. Because $U$ is open, it contains a ball $B(y, \delta)$. Choose $n$ so large that $d(x_n, y) < \delta/2$ and $\varepsilon_n < \delta/2$. Then $B(x_n, \varepsilon_n) \subset B(y, \delta)$, for $d(x, x_n) < \varepsilon_n$ implies $d(x, y) \leq d(x, x_n) + d(x_n, y) < \delta$. Therefore $B(x_n, \varepsilon_n)$ admits a finite subcovering, namely by the single set $U$. This is a contradiction, and we conclude that $S$ has the Heine-Borel property. □

**Corollary (Heine-Borel Property)** A subset of $\mathbb{R}$ or $\mathbb{C}$ is compact if and only if it is closed and bounded.

We have already mentioned this particular consequence. In one direction the conclusion is immediate: We know that a eompact set is bounded and complete; but $\mathbb{R}$ and $\mathbb{C}$ are complete, and complete subsets of a complete space are closed. For the opposite conclusion we need to show that every bounded set in $\mathbb{R}$ or $\mathbb{C}$ is totally bounded. Let us take the case of $\mathbb{C}$. If $X$ is bounded it is contained in a disk, and hence in a square. The square can be subdivided into a finite number of squares with arbitrarily small side, and the squares can in turn be covered by disks with arbitrarily small radius. This proves that $X$ is totally bounded, except for a small point that should not be glossed over. When Definition 7 is applied to a subset $X \subset S$ it is slightly ambiguous, for it is not clear whether the $\varepsilon$-neighborhoods should be with respect to $X$ or with respect to $S$; that is, it is not clear whether we require their centers to lie on $X$. It happens that this is of no avail. In fact, suppose that we have covered $X$ by $\varepsilon$-neighborhoods whose centers do not necessarily lie on $X$. If such a neighborhood does not meet $X$ it is superfluous, and can be dropped. If it does contain a point from $X$, then we can replace it by a $2\varepsilon$-neighborhood around that point, and we obtain a finite covering by $2\varepsilon$-neighborhoods with centers on $X$. For this reason the ambiguity is only apparent, and our proof that bounded subsets of $\mathbb{C}$ are totally bounded is valid.

There is a third characterization of compact sets. It deals with the notion of *limit point* (sometimes called *cluster value*): We say that $y$ is a limit point of the sequence $\{x_n\}$ if there exists a subsequence $\{x_{n_k}\}$ that converges to $y$. A limit point is almost the same as an *accumulation point* of the set formed by the points $x_n$, except that a sequence permits repetitions of the same point. If $y$ is a limit point, every neighborhood of $y$ contains infinitely many $x_n$. The converse is also true. Indeed, suppose that $\varepsilon_k \to 0$. If every $B(y, \varepsilon_k)$ contains infinitely many $x_n$ we can choose subscripts $n_k$, by induction, in such a way that $x_{n_k} \in B(y, \varepsilon_k)$ and $n_{k+1} > n_k$. It is clear that $\{x_{n_k}\}$ converges to $y$.

**Theorem 7. (Bolzano-Weierstrass Theorem)** A metric space is compact if and only if every infinite sequence has a limit point.

The original formulation was that every bounded sequence of complex numbers has a convergent subsequence. It came to be recognized as an important theorem precisely because of the role it plays in the theory of analytic functions.

The first part of the proof is a repetition of an earlier argument. If $y$ is not a limit point of $\{x_n\}$ it has a neighborhood which contains only finitely many $x_n$ (abbreviated version of the correct phrase). If there were no limit points the open sets containing only finitely many $x_n$ would form an open covering. In the compact case we could select a finite subcovering, and it would follow that the sequence is finite. The previous time we used this reasoning was to prove that a compact space is complete. We showed in essence that every sequence has a limit point, and then we observed that a Cauchy sequence with a limit point is necessarily convergent. For strict economy of thought it would thus have been better to prove Theorem 7 before Theorem 6, but we preferred to emphasize the importance of total boundedness as early as possible.

It remains to prove the converse. In the first place it is clear that the Bolzano-Weierstrass property implies completeness. Indeed, we just pointed out that a Cauchy sequence with a limit point must be convergent. Suppose now that the space is not totally bounded. Then there exists an $\varepsilon > 0$ such that the space cannot be covered by finitely many $\varepsilon$-neighborhoods. We construct a sequence $\{x_n\}$ as follows: $x_1$ is arbitrary, and when $x_1, \ldots, x_n$ have been selected we choose $x_{n+1}$ so that it does not lie in $B(x_1, \varepsilon) \cup \cdots \cup B(x_n, \varepsilon)$. This is always possible because these neighborhoods do not cover the whole space. But it is clear that $\{x_n\}$ has no convergent subsequence, for $d(x_m, x_n) > \varepsilon$ for all $m$ and $n$. We conclude that the Bolzano-Weierstrass property implies total boundedness. In view of Theorem 6 that is what we had to prove.

The reader should reflect on the fact that we have exhibited three characterizations of compactness whose logical equivalence is not at all trivial. It should be clear that results of this kind are particularly valuable for the purpose of presenting proofs as concisely as possible.

### 1.5 Continuous Functions

We shall consider functions $f$ which are defined on a metric space $S$ and have values in another metric space $S'$. Functions are also referred to as *mappings*: we say that $f$ maps $S$ into $S'$, and we write $f : S \to S'$. Naturally, we shall be mainly concerned with real- or complex-valued functions; occasionally the latter are allowed to take values in the extended complex plane, ordinary distance being replaced by distance on the Riemann sphere. The space $S$ is the *domain* of the function. We are of course free to consider functions $f$ whose domain is only a subset of $S$, in which rase the domain is regarded as a subspace. In most cases it is safe to slur over the distinction: a function on $S$ and its restriction to a subset are usually denoted by the same symbol. If $X \subset S$ the set of all values $f(x)$ for $x \in S$ is called the *image* of $X$ under $f$, and it is denoted by $f(X)$. The *inverse image* $f^{-1}(X')$ of $X' \subset S'$ consists of all $x \in S$ such that $f(x) \in X^{f\prime}$. Observe that $f\left(f^{-1}(X')\right) \subset X'$, and $f^{-1}\left(f(X)\right) \supset X$.

The definition of a continuous function needs practically no modification: $f$ is *continuous at a* if to every $\varepsilon > 0$ there exists $\delta > 0$ such that $d(x,\, a) < \delta$ implies $d'(f(x),\, f(a)) < \varepsilon$. We are mainly concerned with functions that are continuous at all points in the domain of definition.

The following characterizations are immediate consequences of the definition:

**Definition** Let $(S,\, d)$ and $(S',\, d')$ be metric spaces and $f : S \to S'$ be a function from $S$ into $S'$. Then $f$ is *continuous* if and only if the inverse image of every open set is open. Equivalently, $f$ is continuous if and only if the inverse image of every closed set is closed.

If $f$ is not defined on all of $S$, the words "open" and "closed", when referring to the inverse image, should of course be interpreted relatively to the domain of $f$. It is very important to observe that these properties hold only for the inverse image, not for the direct image. For instance the mapping $f(x) = x^2/\left(1+x^2\right)$ of $\mathbb{R}$ into $\mathbb{R}$ has the image $f\left(\mathbb{R}\right) = \{y \mid 0 \le y < 1\}$ which is neither open nor closed. In this example $f\left(\mathbb{R}\right)$ fails to be closed because $\mathbb{R}$ is not compact. In fact, the following is true:

**Theorem** 8. Under a continuous mapping the image of every compact set is compact, and consequently closed.

Suppose that $f$ is defined and continuous on the compact set $X$. Consider a covering of $f(X)$ by open sets $U$. The inverse images $f^{-1}(U)$ are open and form a covering of $X$. Because $X$ is compact we can select a finite subcovering: $X \subset f^{-1}(U_1) \cup \cdots f^{-1}(U_m)$. It follows that $f(X) \subset U_1 \cup \cdots U_m$ and we have proved that $f(X)$ is compact.

**Corollary** A continuous real-valued function on a compact set has a maximum and a minimum.

The image is a closed bounded subset of $\mathbb{R}$. The existence of a maximum and a minimum follows by Theorem 2.

**Theorem** 9. Under a continuous mapping the image of any connected set is connected.

We may assume that $f$ is defined and continuous on the whole space $S$, and that $f(S)$ is all of $S'$. Suppose that $S' = A \cup B$where $A$ and $B$ are open and disjoint. Then $S = f^{-1}(A) \cup f^{-1}(B)$ is a representation of $S$ as a union of disjoint open sets. If $S$ is connected either $f^{-1}(A) = \emptyset$ or $f^{-1}(B) = \emptyset$, and hence $A = \emptyset$ or $B = \emptyset$. We conclude that $S'$ is connected.

A typical application is the assertion that a real-valued function which is continuous and never zero on a connected set is either always positive or always negative. In fact, the image is connected, and hence an interval. But an interval which contains positive and negative numbers also contains zero.

**Definition** A mapping $f : S \to S'$ is said to be *one to one* if $f(x) = f(y)$ only for $x = y$; it is said to be *onto* if $f(S) = S'$. A mapping with both these properties has an *inverse* $f^{-1}$ defined on $S'$; it satisfies $f^{-1}\left(f(x)\right) = x$ for all $x \in S$ and $f\left(f^{-1}(x')\right) = x'$ for all $x' \in S'$. In this situation, if $f$ and $f^{-1}$ are both continuous we say that $f$is a *topological mapping or a homeomorphism*.

**Definition** Let $(S,\, d)$ and $(S',\, d')$ be metric spaces, $f : S \to S'$ be a mapping from $S$ into $S'$ and $X$ is a subset of $S$. Then $f$ is said to be *uniformly continuous* on $X$ if, to every $\varepsilon > 0$, there exists a $\delta > 0$ such that $d'(f(x_1),\, f(x_2)) < \varepsilon$ for all pairs $(x_1,\, x_2)$ with $d(x_1,\, x_2) < \delta$.

**Theorem 10.** On a compact set every continuous function is uniformly continuous.

The proof is typical of the way the Heine-Borel property can be used. Suppose that $f$ is continuous on a compact set $X$. For every $y \in X$ there is a ball $B(y, \rho)$ such that $d'(f(x), f(y)) < \varepsilon/2$ for $x \in B(y, \rho)$; here $\rho$ may depend on $y$. Consider the covering of $X$ by the smaller balls $B(y, \rho/2)$. There exists a finite subcovering: $X \subset B(y_1, \rho_1/2) \cup \cdots \cup B(y_m, \rho_m/2)$. Let $\delta$ be the smallest of the numbers $\rho_1/2, \ldots, \rho_m/2$, and suppose that $d(x_1, x_2) < \delta$. There is a $y_k$ with $d(x_1, y_k) < \rho_k/2$, and we obtain $d(x_2, y_k) < \rho_k/2 + \delta \leq \rho_k$. Hence $d'(f(x_1), f(y_k)) < \varepsilon/2$ and $d'(f(x_2), f(y_k)) < \varepsilon/2$ so that $d'(f(x_1), f(x_2)) < \varepsilon$ as desired.

On sets which are not compact some continuous functions are uniformly continuous and others are not. For instance, the function $f(z) = z$ is uniformly continuous on the whole complex plane, but the function $f(z) = z^2$ is not.

## 1.6 Topological Spaces

It is not necessary, and not always convenient, to express nearness in terms of distance. The observant reader will have noticed that most results in the preceding sections were formulated in terms of open sets. True enough, we used distances to define open sets, but there is really no strong reason to do this. If we decide to consider the open sets as the primary objects we must postulate axioms that they have to satisfy. The following axioms lead to the commonly accepted definition of a topological space:

**Definition 8.** A *topological space* is a set $T$ together with a collection of its subsets, called *open sets*. The following conditions have to be fulfilled:

(i) The empty set $\emptyset$ and the whole space $T$ are open sets.
(ii) The intersection of any two open sets is an open set.
(iii) The union of an arbitrary collection of open sets is an open set.

We recognize at once that this terminology is consistent with our earlier definition of an open subset of a metric space. Indeed, properties (*ii*) and (*iii*) were strongly emphasized, and (*i*) is trivial.

Closed sets are the complements of open sets, and it is immediately clear how to define interior, closure, boundary, and so on. Neighborhoods could be avoided, but they are rather convenient: $N$ is a neighborhood of $x$ if there exists an open set $U$ such that $x \in U$ and $U \subset N$.

Connectedness was defined purely by means of open sets. Hence the definition carries over to topological spaces, and the theorems remain true. The Heine-Borel property is also one that deals only with open sets. Therefore it makes perfect sense to speak of a compact topological space. However, Theorem 6 becomes meaningless, and Theorem 7 becomes false.

As a matter of fact, the first serious difficulty we encounter is with convergent sequences. The definition is clear: we say that $x_n \to x$ if every neighborhood of $x$ contains all but a finite number of the $x_n$. But if $x_n \to x$ and $x_n \to y$ we are not able to prove that $x = y$. This awkward situation is remedied by introducing a new axiom which characterizes the topological space as a Hausdorff space:

**Definition 9.** A topological space is called a Hausdorff space if any two distinct points are contained in disjoint open sets.

In other words, if $x \neq y$ we require the existence of open sets $U$, $V$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. In the presence of this condition it is obvious that the limit of a convergent sequence is unique. We shall never in this book have occasion to consider a space that is not a Hausdorff space.