Chapter 1

Basic concepts in topology

Concepts like convergence of sequences of points and continuity of functions are fundamental in the subject of mathematical analysis. A more profound study of these concepts, among other things, takes place in the mathematical discipline called general, or pointset, *topology*. In this chapter we shall develop the basic elements of general topology. In particular, we find the proper setting for the following fundamental theorems from classical analysis:

Theorem A. *A continuous real-valued function $f : [a,b] \to \mathbb{R}$ defined in a closed and bounded interval $[a,b]$ is bounded.*

Theorem B. *A continuous real-valued function $f : [a,b] \to \mathbb{R}$ defined in a closed and bounded interval $[a,b]$ attains every value between $f(a)$ and $f(b)$.*

Gradually building up the underlying ideas of the concept of continuity, we first introduce metric spaces and then topological spaces. Theorem A and Theorem B lead respectively to compact sets and connected sets in topological spaces.

1.1 The classical setting for continuity

Informally speaking, a real-valued function $f = f(x) : \mathbb{R} \to \mathbb{R}$ of one real variable $x \in \mathbb{R}$ is said to be continuous if small variations in x only cause small variations in $f(x)$. If we think of x as the input and $f(x)$ as the corresponding output, then this heuristic definition of continuity expresses that small variations in the input only give small variations in the output. This definition, however, is too imprecise to work with. Indeed, consider e.g. $f(x) = 10^{72}x$. This is a linear function, thus in particular a continuous function, but with some justification one can assert that even small variations in x cause large variations in $f(x)$. Hence we need a more precise definition which better encapsulates the idea that we have control over the output of a continuous function.

Definition 1.1.1. Let $U \subseteq \mathbb{R}$ be a subset of $\mathbb{R}$, and let $f : U \to \mathbb{R}$ be a real-valued function. We say that $f : U \to \mathbb{R}$ is *continuous at a point* $x_0 \in U$, provided that

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall x \in U : |x - x_0| < \delta \;\Rightarrow\; |f(x) - f(x_0)| < \varepsilon \,.$$

If $f : U \to \mathbb{R}$ is continuous at every point $x_0 \in U$, we say that f is *continuous*.

Remark 1.1.2. In the above definition we can think of $\varepsilon > 0$ as an admissible margin of error in the output and of $\delta > 0$ as an associated margin of tolerance in the input. Continuity of $f : U \to \mathbb{R}$ at a point $x_0 \in U$ then means that for any given admissible margin of error $\varepsilon > 0$ in the output, we can find a corresponding admissible margin of tolerance $\delta > 0$ in the input such that, as long as x stays within the tolerance $\delta > 0$ of x_0, the actual value $f(x)$ stays within the admissible margin of error $\varepsilon > 0$ of $f(x_0)$.

Continuity of $f : U \to \mathbb{R}$ means that f is continuous at every point $x_0 \in U$. Thus, in Definition 1.1.1, a $\delta > 0$ corresponding to a given $\varepsilon > 0$ in general depends on the point $x_0 \in U$. By using quantifiers, the definition of continuity of $f : U \to \mathbb{R}$ may be written

$$\forall x_0 \in U \;\; \forall \varepsilon > 0 \;\; \exists \delta > 0 \; \forall x \in U : \; |x - x_0| < \delta \;\Rightarrow\; |f(x) - f(x_0)| < \varepsilon \,.$$

This should not be confused with

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall x, y \in U \; : \; |x - y| < \delta \;\Rightarrow\; |f(x) - f(y)| < \varepsilon \,.$$

In the second case, the tolerance $\delta > 0$ does not depend on y $(= x_0)$, whereas it does in the first case. In fact, in the second case we define that $f : U \to \mathbb{R}$ is *uniformly continuous* .

Example 1.1.3. Consider $f(x) = x^2$ for $x \in \mathbb{R}$. For all $x, y \in \mathbb{R}$ we have

$$|f(x) - f(y)| = \left|x^2 - y^2\right| = |x + y| \cdot |x - y| \,.$$

For any $\varepsilon > 0$ and any fixed $y \in \mathbb{R}$, we can find $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ when $|x - y| < \delta$. Thus the function f is continuous. However, we cannot choose $\delta > 0$ independently of y, since by choosing x and y sufficiently large, we can obtain that $|f(x) - f(y)| > \varepsilon$ even though $|x - y| < \delta$. Hence, the function f is not uniformly continuous.

Remark 1.1.4. Replacing the strict inequalities $< \delta$ and $< \varepsilon$ by weak inequalities $\leq \delta$ and $\leq \varepsilon$ does not affect the definitions.

Let $\mathbb{R}^n$ denote the space consisting of n-tuples $x = (x_1, \ldots, x_n)$ of real numbers, and let $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ be k real-valued functions of n variables. As usual, we can collect these into a mapping

$$f = (f_1, \ldots, f_k) : \mathbb{R}^n \to \mathbb{R}^k,$$

where $f(x) = (f_1(x), \ldots, f_k(x))$, which we can elaborate to

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_k(x_1, \ldots, x_n)).$$

If $U \subseteq \mathbb{R}^n$ is a subset of $\mathbb{R}^n$, and the functions $f_1, \ldots, f_k$ are defined only in U, we obtain a mapping

$$f = (f_1, \ldots, f_k) : U \to \mathbb{R}^k.$$

The above definition of continuity of the function $f : \mathbb{R} \to \mathbb{R}$, or more generally, $f : U \to \mathbb{R}$ defined in a subset $U \subseteq \mathbb{R}$, can easily be generalized to concern also mappings $f : \mathbb{R}^n \to \mathbb{R}^k$, or more generally, $f : U \to \mathbb{R}^k$ defined in a subset $U \subseteq \mathbb{R}^n$. We only have to replace $|x - x_0|$, respectively $|f(x) - f(x_0)|$, by the Euclidean distances in $\mathbb{R}^n$, respectively $\mathbb{R}^k$, formally defined in Example 1.2.4; cf. Figure 1.1 for $n = k = 2$.

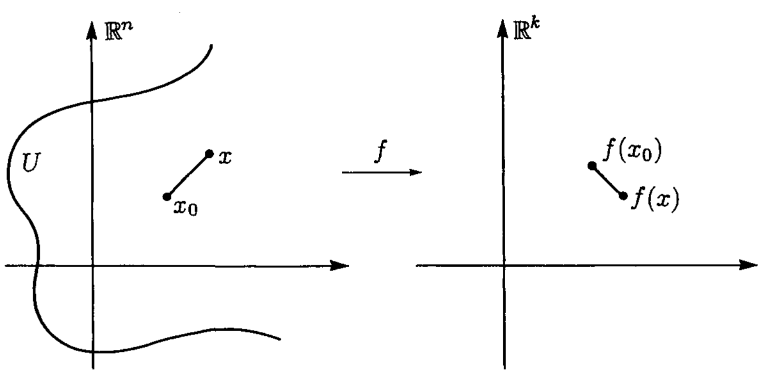


Figure 1.1

1.2 Metric spaces

It is useful to introduce a notion of continuity of mappings in more abstract situations, where input for the mappings are not merely n-tuples of real numbers, but perhaps a collection of functions or a collection of differential equations. We therefore want to generalize the definition of continuity so that it can be applied to mappings $f : X \to Y$, where X and Y are more general spaces than just $\mathbb{R}^n$ and $\mathbb{R}^k$ or subsets of these spaces. We can perform an immediate generalization, which can be used in many situations, when X and Y are equipped with the notion of a distance.

A distance function (or a metric) on a set S first of all has to be a function $d : S \times S \to \mathbb{R}$ which to any pair of points $x, y \in S$ associates a real number $d(x, y)$, called the *distance* from x to y. Furthermore, to get a reasonable notion of distance, it has proved fruitful to require that the following three conditions are satisfied:

MET 1 (positive definite)

$$d(x,y) \geq 0, \quad \text{for all} \quad x, y \in S, \qquad = 0 \iff x = y.$$

MET 2 (symmetry)

$$d(x,y) = d(y,x), \quad \text{for all} \quad x, y \in S.$$

MET 3 (the triangle inequality)

$$d(x,z) \leq d(x,y) + d(y,z), \quad \text{for all} \quad x, y, z \in S.$$

Definition 1.2.1. A function $d : S \times S \to \mathbb{R}$, which satisfies MET 1, MET 2, and MET 3 is called a *metric* on S. A set S together with a specific metric on S is called a *metric space*.

Remark 1.2.2. A metric space is, in other words, a pair (S, d) consisting of a set S and a metric d on S. The same set S can be equipped with several different metrics, in which case we also consider the corresponding metric spaces as different. When the context leaves no doubt as to which metric is being considered, usually the metric is not mentioned explicitly.

The idea of metric spaces was introduced by the French mathematician Maurice Fréchet (1878–1973) in his doctoral thesis of 1906.

Let $\mathbb{R}_0^+$ denote the set of non-negative real numbers. Condition MET 1 then states that a metric is actually a function $d : S \times S \to \mathbb{R}_0^+$.

We now give some examples of metric spaces.

Example 1.2.3. Consider $S = \mathbb{R}$. For the real numbers $x, y \in \mathbb{R}$ we set

$$d(x,y) = |x - y| \ .$$

Clearly, the function d so defined satisfies MET 1 and MET 2. The triangle inequality, MET 3, follows this way:

$$\begin{aligned} d(x,z) = |x-z| &= |x-y+y-z| \\ &\le |x-y| + |y-z| = d(x,y) + d(y,z). \end{aligned}$$

As we shall see later, the above example is typical for a metric space arising from a normed vector space.

Example 1.2.4. Consider $S = \mathbb{R}^n$.

For points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ we set

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \ .$$

Again, MET 1 and MET 2 follow easily. To prove that MET 3 is satisfied we need to make some preparations.

The space $\mathbb{R}^n$ is an n-dimensional vector space with the usual coordinate-wise defined laws of addition and multiplication by scalars. The points in $\mathbb{R}^n$ are then identified with vectors. We define the *inner product* by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i \ ,$$

and the associated *norm* by

$$||x|| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} {x_i}^2} \ .$$

Then we have the *Cauchy-Schwarz inequality:*

$$|\langle x, y \rangle| \le ||x|| \cdot ||y|| \ .$$

Proof. The Cauchy-Schwarz inequality is satisfied in any vector space equipped with an inner product and the proof we present is completely general.

Let $x, y \in \mathbb{R}^n$ be arbitrarily chosen, but fixed vectors in $\mathbb{R}^n$. For every real number $t \in \mathbb{R}$ we have the inequality

$$0 \le \langle x + ty, x + ty \rangle = ||y||^2 t^2 + 2\langle x, y \rangle t + ||x||^2 \ .$$

For $y \neq 0$, this describes a polynomial of degree 2 in t (a parabola) with at most one zero. Therefore, the discriminant of the polynomial must satisfy

$$4\langle x, y \rangle^2 - 4||x||^2 ||y||^2 \le 0 \ .$$

From this we get the inequality $|\langle x, y \rangle| \le ||x|| \cdot ||y||$ as asserted.

For $y = 0$, the inequality is trivially satisfied. □

The Cauchy-Schwarz inequality implies the *triangle inequality:*

$$||x+y|| \leq ||x|| + ||y||.$$

Proof. Using the Cauchy-Schwarz inequality, the small computation

$$\begin{aligned} ||x+y||^2 &= \langle x+y, x+y\rangle = ||x||^2 + ||y||^2 + 2\langle x,y\rangle \\ &\leq ||x||^2 + ||y||^2 + 2||x|| \cdot ||y|| \\ &= (||x|| + ||y||)^2 \end{aligned}$$

reveals the triangle inequality. □

By noting that

$$d(x,y) = ||x-y||,$$

the condition MET 3 now follows precisely as in Example 1.2.3:

$$\begin{aligned} d(x,z) = ||x-z|| &= ||x-y+y-z|| \\ &\leq ||x-y|| + ||y-z|| = d(x,y) + d(y,z). \end{aligned}$$

Thereby we have verified that $\mathbb{R}^n$ equipped with the distance function d is a metric space. This metric space is called n-dimensional *Euclidean space*, and the metric d, *the Euclidean metric* on $\mathbb{R}^n$.

For $n = 2, 3$ we rediscover the usual concept of distance in the plane, respectively 3-dimensional space, when these are identified with $\mathbb{R}^2$, respectively $\mathbb{R}^3$, by choosing a standard orthogonal coordinate system.

Before the next example, we remind the reader about a property of the real numbers, namely the existence of a supremum (and infimum). Together with the well-known algebraic properties, the existence of a supremum characterizes the system of real numbers.

Property (Supremum). *Every non-empty subset A of $\mathbb{R}$ which is bounded from above has a smallest majorant, which is called the* least upper bound *of A, or* supremum *of A, and is denoted by* $\sup A$.

From the existence of a supremum we easily get:

Property (Infimum). *Every non-empty subset A of $\mathbb{R}$ which is bounded from below has a greatest minorant, which is called the* greatest lower bound *of A, or* infimum *of A, and is denoted by* $\inf A$.

Example 1.2.5. Let $[a,b]$ be a closed and bounded interval of $\mathbb{R}$, and consider the set S of all bounded real-valued functions $f : [a,b] \to \mathbb{R}$ defined in $[a,b]$.

Actually, the construction below holds for the set of bounded functions defined on an arbitrary set K.

For $f, g \in S$ we put

$$\begin{aligned} d(f,g) &= \sup_{a \le x \le b} |f(x) - g(x)| \\ &= \sup \left\{ |f(x) - g(x)| \;\middle|\; a \le x \le b \right\}. \end{aligned}$$

Note that since

$$A = \left\{ |f(x) - g(x)| \;\middle|\; a \le x \le b \right\}$$

is bounded from above, the number $d(f,g) = \sup A$ exists. At the same time we have indicated different ways of writing $\sup A$ in the present example.

Since $0 \le |f(x) - g(x)| \le \sup A$ for all $a \le x \le b$, MET 1 follows easily.

Also MET 2 is trivially satisfied, since $|f(x) - g(x)| = |g(x) - f(x)|$ for all $a \le x \le b$.

To prove that MET 3 is satisfied we proceed as follows.

Let f, g, h be three functions in S. For every $a \le x \le b$, we then have

$$\begin{aligned} |f(x) - h(x)| &= |f(x) - g(x) + g(x) - h(x)| \\ &\le |f(x) - g(x)| + |g(x) - h(x)| \\ &\le \sup_{a \le x \le b} |f(x) - g(x)| + \sup_{a \le x \le b} |g(x) - h(x)| \\ &= d(f,g) + d(g,h). \end{aligned}$$

From this we see that $d(f,g) + d(g,h)$ is an upper bound for

$$\left\{ |f(x) - h(x)| \;\middle|\; a \le x \le b \right\}.$$

Since any upper bound is greater than or equal to the least upper bound, we conclude that

$$d(f,h) = \sup_{a \le x \le b} |f(x) - h(x)| \le d(f,g) + d(g,h).$$

Hence d is a metric on S.

If we define addition of functions in S and multiplication of functions by real numbers using the obvious pointwise defined operations, S becomes a real vector space. This vector space has infinite dimension, since there is no system of finitely many functions spanning S.

Example 1.2.6. Let S be any set. Set

$$d(x,y) = \begin{cases} 1 & \text{for } x \neq y \\ 0 & \text{for } x = y. \end{cases}$$

It is easy to see that d is a metric on S. This metric on S is called *the discrete metric*.

Example 1.2.7. Let (S,d) be a metric space, and let $T \subseteq S$ be a subset of S. Then T inherits a metric from (S,d) called *the induced metric* . More formally: If $d : S \times S \to \mathbb{R}_0^+$ is the metric on S, we get the induced metric on T by taking the restriction of d to $T \times T$.

We finish this section with the formal definition of the notion of continuity of a mapping within the setting of metric spaces.

Definition 1.2.8. Let (X,d_X) and (Y,d_Y) be metric spaces. We say that the mapping $f : X \to Y$ is *continuous at a point* $x_0 \in X$, if

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 \ \ \forall x \in X : d_X(x,x_0) < \delta \quad \Rightarrow \quad d_Y(f(x),f(x_0)) < \varepsilon.$$

We say that the mapping $f : X \to Y$ is *continuous*, if f is continuous at every point $x_0 \in X$.

1.3 The topology of a metric space

The above definition of continuity of a mapping between metric spaces is strongly dependent on the metrics chosen. On the other hand it is clear that the dependence on certain specific metrics should not be too strong, since, for example, we will clearly allow multiplication of a metric by a fixed universal constant corresponding to a change of the unit length. A closer study of continuity and its independence of the choice of certain specific metrics naturally leads to the concept of *topology.*

We start by reformulating the definition of continuity of a mapping $f : X \to Y$ between metric spaces (X,d_X) and (Y,d_Y).

First, a general definition. Let (S,d) be an arbitrary metric space, let x_0 be a point in S, and let $r \in \mathbb{R}^+$ be a positive real number. Then we call

$$B_r(x_0) = \{x \in S \mid d(x_0,x) < r\}$$

the *open ball* or the *open sphere* in S with *centre* x_0 and *radius* r.

The definition of continuity of the mapping $f : X \to Y$ at a point $x_0 \in X$ (Definition 1.2.8) can now be formulated as follows:

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 : \ f\left(B_\delta(x_0)\right) \subseteq B_\varepsilon\left(f(x_0)\right),$$

or equivalently:

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 : B_\delta(x_0) \subseteq f^{-1}\left(B_\varepsilon\left(f(x_0)\right)\right).$$

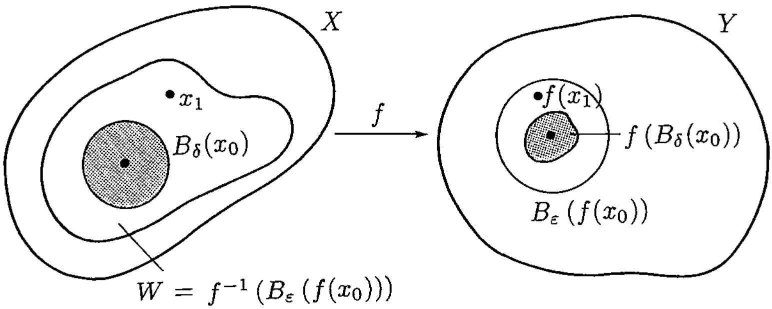


Figure 1.2

Now assume that $f : X \to Y$ is continuous, and consider the subset $W = f^{-1}(B_\varepsilon(f(x_0)))$ in X; cf. Figure 1.2. From the definition of W it is clear that $B_\delta(x_0) \subseteq W$. Thus the point $x_0 \in W$ has the property that there exists an open ball with centre in x_0, completely contained in W. Now, take any point $x_1 \in W$. We will show that continuity of f implies that the point x_1 has the same property as x_0. First note that $f(x_1) \in B_\varepsilon(f(x_0))$. Then choose an ε_1 with $0 < \varepsilon_1 \leq \varepsilon - d_Y(f(x_1), f(x_0))$. Since $f : X \to Y$ is continuous at x_1, we can, for $\varepsilon_1 > 0$, choose a $\delta_1 > 0$, so that $f(B_{\delta_1}(x_1)) \subseteq B_{\varepsilon_1}(f(x_1))$.

Assertion 1.3.1. $B_{\delta_1}(x_1) \subseteq W$.

Proof. We need only show that $B_{\varepsilon_1}(f(x_1)) \subseteq B_\varepsilon(f(x_0))$. This follows from the triangle inequality, since for $y \in B_{\varepsilon_1}(f(x_1))$ we have

$$\begin{aligned} d_Y(f(x_0), y) &\leq d_Y(f(x_0), f(x_1)) + d_Y(f(x_1), y) \\ &< d_Y(f(x_0), f(x_1)) + \varepsilon_1 \leq \varepsilon, \end{aligned}$$

which shows that $y \in B_\varepsilon(f(x_0))$. □

Hence every point in W is the centre of an open ball completely contained in W. This motivates the following definition.

Definition 1.3.2. Let (S, d) be a metric space. A subset W of S is called an *open set* in the metric space (S, d), if for every $x \in W$ there exists a $\delta > 0$ such that $B_\delta(x) \subseteq W$.

Remark 1.3.3. An open ball $B_r(x_0)$ in the metric space (S, d) is an open set in S, since for $x \in B_r(x_0)$ the triangle inequality shows that $B_\delta(x) \subseteq B_r(x_0)$, if $0 < \delta \leq r - d(x_0, x)$.

Example 1.3.4. Open intervals in $\mathbb{R}$ are open sets.

In $\mathbb{R}^2$, the sets

$$\{(x,y) \mid -2 < x < 3, -1 < y < 1\} \quad \text{and} \quad \{(x,y) \mid x^2 + y^2 \neq 1\}$$

are examples of open sets.

A subset of $\mathbb{R}^n$ described by a finite number of strict inequalities defined by continuous functions is open.

As the following theorem shows, we can describe continuity of a mapping in a very simple way, using open sets.

Theorem 1.3.5. *Let $f : X \to Y$ be a mapping between metric spaces (X, d_X) and (Y, d_Y). Then f is continuous if and only if for every open set V in Y, the set $f^{-1}(V)$ is an open set in X.*

Proof. First assume that f is continuous. Let V be an arbitrary open set in Y. We have to show that $f^{-1}(V)$ is an open set in X. For that purpose let $x \in f^{-1}(V)$ be an arbitrary point in $f^{-1}(V)$ in case $f^{-1}(V) \neq \emptyset$. Since $f(x) \in V$, and V is open in Y, we can find an $\varepsilon > 0$ such that $B_\varepsilon(f(x)) \subseteq V$. Since f is continuous at x, for this $\varepsilon > 0$ we can find a $\delta > 0$, so that $B_\delta(x) \subseteq f^{-1}(B_\varepsilon(f(x)))$. But then $B_\delta(x) \subseteq f^{-1}(V)$, and since $x \in f^{-1}(V)$ was an arbitrarily chosen point, this shows that $f^{-1}(V)$ is an open set in X, possibly the empty set.

Assume now conversely, that all preimages of open sets in Y under f are open sets in X. We then have to show that f is continuous at every point $x \in X$. Therefore let $x_0 \in X$ be an arbitrary point in X, and let $\varepsilon > 0$ be given. According to Remark 1.3.3, $B_\varepsilon(f(x_0))$ is an open set in Y, and therefore by the assumption, $f^{-1}(B_\varepsilon(f(x_0)))$ is an open set in X. But then there exists a $\delta > 0$, such that $B_\delta(x_0) \subseteq f^{-1}(B_\varepsilon(f(x_0)))$. This shows that f is continuous at x_0. □

Theorem 1.3.5 shows that open sets are decisive for the study of continuity of mappings between metric spaces, rather than the metrics themselves. In particular, if two different metrics on the same set give rise to the same family of open sets, then any mapping which is continuous with respect to one of the metrics is also continuous with respect to the other metric.

The family of open sets is called the *topology* in the metric space.

1.4 Topological spaces

We just demonstrated that all we need in order to define the concept of continuity of a mapping $f : X \to Y$ is a topology on each of the sets X and Y. In this section we formally define what we understand by a topology on a set.

Let S be a fixed set. A topology on S first of all has to be a family $\mathcal{T}$ of subsets of S. However, it is clear that we cannot choose just any family of subsets of S if we want to get something sensible out of it. It turns out to be reasonable to require that the family of subsets $\mathcal{T}$ in S satisfies the following conditions:

TOP 1 If $\{U_i \in \mathcal{T} \mid i \in I\}$ is an arbitrary system of subsets in S from $\mathcal{T}$, then the union $\cup\{U_i \in \mathcal{T} \mid i \in I\}$ of these subsets also belongs to $\mathcal{T}$.

TOP 2 If $U_1, \ldots, U_k$ is an arbitrary finite system of subsets in S from $\mathcal{T}$, then the intersection $\cap_{i=1}^{k} U_i$ of these subsets also belongs to $\mathcal{T}$.

TOP 3 The empty set $\emptyset$, and the set S itself belong to $\mathcal{T}$.

It is a relatively simple exercise to show that the family of open sets in a metric space satisfies the conditions TOP 1, TOP 2, and TOP 3. So the following definition is in accordance with the terminology introduced in the setting of metric spaces:

Definition 1.4.1. A family of subsets $\mathcal{T}$ in the set S satisfying the conditions TOP 1, TOP 2, and TOP 3, is called a *topology* on S. A set S together with a specific topology $\mathcal{T}$ on S is called a *topological space*. The specified subsets in $\mathcal{T}$ are called the *open sets* in the topological space.

Remark 1.4.2. A topological space is in other words a pair $(S, \mathcal{T})$ consisting of a set S and a topology $\mathcal{T}$ on S. Usually, however, we only talk about the topological space S, and let it be implicit that there is given an associated topology $\mathcal{T}$ on S.

A common and shorter formulation of the conditions TOP 1 and TOP 2 is that the family of subsets of the set S in the topology $\mathcal{T}$ is closed under arbitrary unions of sets (TOP 1) and finite intersections of sets (TOP 2) from $\mathcal{T}$.

The restriction to finiteness in TOP 2 is needed. For example, on the real axis $\mathbb{R}$ consider the system of open intervals $]-1/n, 1/n[$ for $n = 1, 2, \ldots$. The intersection of the intervals is the set containing only 0, and this is not an open set in $\mathbb{R}$.

Remark 1.4.3. To test whether a family $\mathcal{T}$ of subsets of a set S satisfies the condition TOP 2, it is clearly sufficient to inspect whether the intersection $U \cap V$ of just two arbitrary subsets U and V from the family $\mathcal{T}$ again belongs to $\mathcal{T}$.

We shall now list some examples of topological spaces.

Example 1.4.4. Any metric on a set S determines a topology on S. It consists of the family $\mathcal{T}$ of open sets corresponding to the metric. Every metric space is then a topological space.

Conversely, not every topological space admits a metric. There are examples of topological spaces which are not *metrizable.*

The topology on $\mathbb{R}^n$ coming from the Euclidean metric, is called *the usual topology*, or *the Euclidean topology*, on $\mathbb{R}^n$. Unless explicitly specified, $\mathbb{R}^n$ is always equipped with this topology.

Example 1.4.5. Let S be an arbitrary set. Then the set of *all* subsets of S will of course satisfy TOP 1, TOP 2, and TOP 3, and thus is a topology on S. This topology is called *the discrete topology* on S, *the largest topology* on S, or *the finest topology* on S.

The discrete topology on S arises from the discrete metric on S. This follows easily by observing that with respect to the discrete metric on S, any subset in S consisting of just one point in S is an open set. (For $0 < \delta \leq 1$ the open ball $B_\delta(x)$ contains nothing but the point x itself.)

Example 1.4.6. If we go to the other extreme, we can take the family of subsets in S, consisting of only the empty set $\emptyset$ and the set S itself. This family of subsets in S forms a topology on S, which we call *the indiscrete topology* on S, *the smallest topology* on S, or *the coarsest topology* on S. When S contains at least two elements, the indiscrete topology on S is not metrizable.

Example 1.4.7. Let S be a topological space with the topology $\mathcal{T}$, and let A be an arbitrary subset in S (not necessarily from $\mathcal{T}$).

Consider the family of subsets in A which we get by taking the intersections of A and the subsets of S from the family $\mathcal{T}$; cf. Figure 1.3. We denote this family of subsets in A by $\mathcal{T}_A$, i.e.

$$\mathcal{T}_A = \{V = A \cap U \mid U \in \mathcal{T}\}.$$

It is easy to verify that $\mathcal{T}_A$ is a topology on A. This topology is called *the induced topology* on A, *the subspace topology* on A, or *the trace topology* on A.

If the topology $\mathcal{T}$ on S stems from a metric on S, then the induced topology $\mathcal{T}_A$ on A stems from the induced metric on A.

Example 1.4.8. Let S and T be topological spaces. We wish to define a topology on the product set $S \times T$.

It is tempting to take sets of the form $U \times V$, where U is open in S and V is open in T, to be the open sets in $S \times T$. But if we do so, then TOP 1 can fail, since indeed a set obtained as the union $(U_1 \times V_1) \cup (U_2 \times V_2)$ does not necessarily have the right product form. (It is easy to construct simple counterexamples in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Consider for instance two rectangles.)

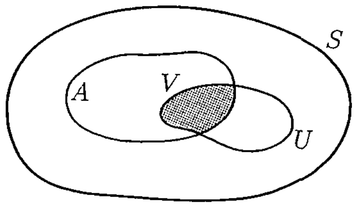


Figure 1.3

To avoid this difficulty, we consider the larger family of subsets in $S \times T$, which arises by including arbitrary unions of sets of the form $U \times V$. Hence, we consider the family $\mathcal{T}_{S\times T}$ of subsets in $S \times T$ of the form

$$\bigcup_{i\in I} U_i \times V_i,$$

where U_i is an open set in S, V_i is an open set in T and I is any index set.

It should come as no surprise that the family $\mathcal{T}_{S\times T}$ is a topology on $S \times T$. The conditions TOP 1 and TOP 3 are now both trivially satisfied. To prove that TOP 2 is satisfied, we just have to use that

$$\Big(\bigcup_{i\in I} U_i \times V_i\Big) \bigcap \Big(\bigcup_{j\in J} U_j' \times V_j'\Big) = \bigcup_{i\in I}\bigcup_{j\in J} (U_i \cap U_j') \times (V_i \cap V_j').$$

The topology $\mathcal{T}_{S\times T}$ is called *the product topology* on $S \times T$.

If the topologies on S and T stem from metrics, then the product topology on $S \times T$ also stems from a metric.

Example 1.4.9. Let S be a topological space, and let $\tilde{S}$ be an arbitrary set. Furthermore, we are given a mapping $\pi : S \to \tilde{S}$. Usually π will be surjective, but this is not necessary for what follows.

Let $\mathcal{T}$ be the topology on S. Consider the matching family $\tilde{\mathcal{T}}$ of subsets V in $\tilde{S}$, for which the corresponding preimages $U = \pi^{-1}(V)$ under π belong to $\mathcal{T}$; cf. Figure 1.4. In other words we set

$$\tilde{\mathcal{T}} = \{V \subseteq \tilde{S} \mid U = \pi^{-1}(V) \in \mathcal{T}\}.$$

By using the set theoretical formulae

$$\pi^{-1}\Big(\bigcup_{i\in I} V_i\Big) = \bigcup_{i\in I} \pi^{-1}(V_i) \quad \text{and} \quad \pi^{-1}\Big(\bigcap_{i\in I} V_i\Big) = \bigcap_{i\in I} \pi^{-1}(V_i),$$

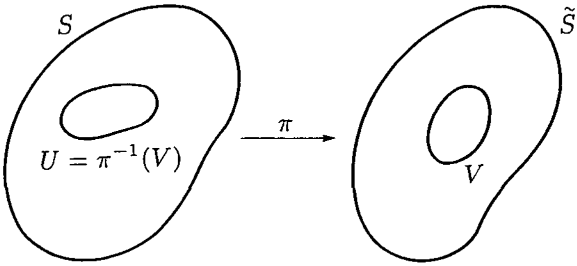


Figure 1.4

it is easily verified that $\tilde{\mathcal{T}}$ is a topology on $\tilde{S}$.

The topology $\tilde{\mathcal{T}}$ on $\tilde{S}$ is called *the quotient topology* on $\tilde{S}$ induced by the mapping π. The name is due to the fact that $\tilde{S}$ often arises from S by a process of identification, under which points in $\tilde{S}$ emerge as equivalence classes of points in S.

Even if the topology on S stems from a metric, the quotient topology on $\tilde{S}$ does not necessarily do so, and even when it does, it is usually not in a natural way.

The original motivation for introducing the notion of a topological space was to study continuity of mappings in a more general setting than metric spaces. In view of Theorem 1.3.5, the only reasonable definition of continuity of a mapping between topological spaces is the following definition.

Definition 1.4.10. Let $f : X \to Y$ be a mapping between topological spaces X and Y. We say that f is *continuous* if, for every open set V in Y, the set $f^{-1}(V)$ is open in X.

As Theorem 1.3.5 shows, Definition 1.4.10 is the immediate generalization of the concept of continuity from the setting of metric spaces to the setting of topological spaces.

What have we gained by this? First of all, we avoid referring to specific metrics. Secondly – and of course most importantly – some topological spaces can only be equipped with a metric in a very unnatural way. In fact, many important topological spaces, like certain function spaces, are not metrizable.

Another benefit is that we now have an elegant tool for handling theoretical questions in connection with continuous functions. Consider for instance

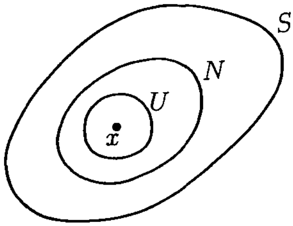


Figure 1.5

the following proof that the composition of two continuous mappings is itself continuous.

Theorem 1.4.11. *If $f : X \to Y$ and $g : Y \to Z$ are continuous mappings between topological spaces, then the* composite mapping $g \circ f : X \to Z$ *is also continuous.*

Proof. Let V be an open set in Z. We need to show that $(g \circ f)^{-1}(V)$ is an open set in X. With this in mind note that

$$(g \circ f)^{-1}(V) = f^{-1}\left(g^{-1}(V)\right).$$

Since g is continuous, $g^{-1}(V)$ is an open set in Y, and since f is continuous, $f^{-1}\left(g^{-1}(V)\right)$ is an open set in X. This completes the proof. □

1.5 Local theory

1.5.1 Neighbourhoods

Let S be a topological space, and let $x \in S$ be a point in S.

Definition 1.5.1. An open set U in S which contains x, is called an *open neighbourhood* of x in S. More generally, a subset N of S which contains x *and* an open neighbourhood U of x, that is $x \in U \subseteq N$ (see Figure 1.5), is called a *neighbourhood* of x in S.

Example 1.5.2. On the real axis $\mathbb{R}$, the closed interval $[-1, 1]$ is a neighbourhood of 0, but not of -1 or 1.

Lemma 1.5.3. *A subset W of S is open if and only if every point $x \in W$ has a neighbourhood N_x in S such that $N_x \subseteq W$.*

Proof. First, assume that W is open in S. Then W itself is an open neighbourhood N_x of every point $x \in W$.

Next, assume that every point $x \in W$ has a neighbourhood N_x with $N_x \subseteq W$. According to the definition of a neighbourhood, every point $x \in W$ then also has an open neighbourhood U_x such that $U_x \subseteq W$. Since $W = \cup_{x \in W} U_x$ it now follows by TOP 1 that W is open in S. □

1.5.2 Continuity at a point

Lemma 1.5.3 reveals that neighbourhoods play the same role for topological spaces as open balls do for metric spaces. As an example we can define continuity of a mapping at a point.

Definition 1.5.4. Let $f : X \to Y$ be a mapping between topological spaces. We say that f is *continuous at a point* $x_0 \in X$, if for every neighbourhood M of $f(x_0)$ in Y there exists a neighbourhood N of x_0 in X, such that $f(N) \subseteq M$.

It is now easy to prove the following theorem.

Theorem 1.5.5. *Let $f : X \to Y$ be a mapping between topological spaces. Then f is continuous if and only if f is continuous at every point $x \in X$.*

Proof. First, assume that f is continuous, and consider a point $x_0 \in X$. Let M be an arbitrary neighbourhood of $f(x_0)$ in Y. Then there exists an open neighbourhood V of $f(x_0)$ in Y such that $V \subseteq M$, and since f is continuous, $N = f^{-1}(V)$ is an open neighbourhood of x_0 in X for which $f(N) \subseteq V \subseteq M$. This proves that f is continuous at the arbitrarily chosen point $x_0 \in X$.

Now, assume conversely that f is continuous at every point $x \in X$, and let V be an arbitrarily chosen open set in Y. We have to prove that $f^{-1}(V)$ is an open set in X. According to Lemma 1.5.3, we only need to show that any point $x \in f^{-1}(V)$ has a neighbourhood N_x in X such that $N_x \subseteq f^{-1}(V)$, or equivalently that $f(N_x) \subseteq V$. However, this is obvious since f is continuous at x, and V is an open neighbourhood of $f(x)$ in Y. □

With Theorem 1.5.5 we have completed the circle of ideas by getting back to the original definition of continuity in the setting of metric spaces.

1.5.3 Fundamental system of neighbourhoods. Basis for a topology

We finish this section with a few remarks about the concept of a *basis* for a topology.

Definition 1.5.6. Let $(S, \mathcal{T})$ be a topological space.

A *fundamental system of neighbourhoods* (or a *local basis*) for a point $x \in S$ is a system of neighbourhoods $\{N_i \mid i \in I\}$ of $x \in S$ with the property that every neighbourhood N of $x \in S$ contains a neighbourhood from the system $\{N_i \mid i \in I\}$.

A *basis* for the topology $\mathcal{T}$ is a system of open sets $\{B_i \mid i \in I\}$ from $\mathcal{T}$ with the property that every open set U from $\mathcal{T}$ can be written as the union of open sets from the system $\{B_i \mid i \in I\}$.

A particularly important case is when the index set I in Definition 1.5.6 is either finite or the set of natural numbers $\mathbb{N}$. In the latter case we say that the corresponding systems of neighbourhoods or open sets are *countable*. In either situation, we say that the systems are *at most countable* or *numerable* (they can be assigned numbers).

Definition 1.5.7. A topological space $(S, \mathcal{T})$ satisfies the *first axiom of countability* if every point $x \in S$ has a numerable fundamental system of neighbourhoods.

A topological space $(S, \mathcal{T})$ satisfies the *second axiom of countability* if the topology $\mathcal{T}$ has a numerable basis.

Example 1.5.8. Let (S, d) be a metric space, and let $x \in S$ be an arbitrary point in S. It is clear that the system of open sets $\{B_{1/n}(x) \mid n \in \mathbb{N}\}$ is a numerable fundamental system of neighbourhoods for $x \in S$. So a metric space satisfies the first axiom of countability. Therefore, in particular, a topological space which does not satisfy the first axiom of countability cannot be metrizable.

A topological space $(S, \mathcal{T})$ satisfying the second axiom of countability clearly also satisfies the first axiom of countability. The converse is not always the case. For instance, an infinite non-countable set (e.g. the real numbers $\mathbb{R}$, cf. Theorem 1.10.4) equipped with the discrete topology satisfies the first but not the second axiom of countability.

1.6 Points in relation to a subset

Let W be an arbitrary subset of the topological space S.

Definition 1.6.1. A point $x \in W$ is called an *interior point* of W if there exists a neighbourhood N_x of x in S completely contained in W, i.e. $N_x \subseteq W$. The collection of interior points of W is called *the interior* of W, and is denoted by $\operatorname{int} W$.

It may happen that int W is the empty set $\emptyset$. For example both the set of rational numbers $\mathbb{Q}$ and the set of irrational numbers $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ have empty interiors in $\mathbb{R}$. The interior of a subset depends on the topological space in which it is considered as a subset. For instance, the interior of the closed interval $[a, b]$ considered as a subset of $\mathbb{R}$ is exactly the open interval $]a, b[$, whereas it has an empty interior considered as a subset of $\mathbb{R}^2$. In some sense, int W measures how 'fat' W is, considered as a subset of S.

Theorem 1.6.2. *The interior of a subset W in the topological space S has the following properties:*

(i) int W *is an open set in* S.

(ii) *If U is an open set in S so that $U \subseteq W$, then $U \subseteq$* int W.
(int W *is the 'largest' open set in S contained in W*).

(iii) *If U is an open set in S so that* int $W \subseteq U \subseteq W$, *then* $U =$ int W.

(iv) int W *is the union of all open sets U in S for which $U \subseteq W$.*

Proof. Proof of (i). If int $W = \emptyset$, the statement is trivially true. So assume that int $W \neq \emptyset$. According to the definition of an interior point, we can, for any $x \in$ int W, find an open neighbourhood U_x of x in S such that $U_x \subseteq W$. Since U_x is an open neighbourhood of any of its points, clearly $U_x \subseteq$ int W. This shows that int W is an open set in S by Lemma 1.5.3.

Proof of (ii). Let U be an open set in S so that $U \subseteq W$. Since U is open, for any $x \in U$ there is a neighbourhood N_x of x in S, such that $N_x \subseteq U \subseteq W$. In fact, U itself is such a neighbourhood. But then $x \in$ int W, proving that $U \subseteq$ int W.

Proof of (iii). Follows immediately from (ii).

Proof of (iv). The union of all open sets U in S such that $U \subseteq W$ is itself an open set $\tilde{U}$ in S such that $\tilde{U} \subseteq W$. According to (i) int W is itself one of the open sets in this union, thus int $W \subseteq \tilde{U} \subseteq W$. But then according to (iii), we have that $\tilde{U} =$ int W. □

Theorem 1.6.2 has the following immediate corollary.

Corollary 1.6.3. *A subset W of a topological space S is open if and only if* $W =$ int W.

Complementary to the concept of an interior point, we have the concept of an exterior point.

Definition 1.6.4. A point $x \in S$ is called an *exterior point* of the subset W in the topological space S, if x is an interior point in the set $S \setminus W$. The collection of exterior points of W is called *the exterior* of W in S.

We do not need any special symbol for the exterior of a set, since *the exterior of W in S* is exactly the interior of $S \setminus W$ in S, i.e. int $(S \setminus W)$.

Finally, we have a third type of points in relation to W considered as a subset of S.

Definition 1.6.5. A point $x \in S$ is called a *boundary point* of the subset W in the topological space S if it is neither an interior point nor an exterior point of W. In other words, a point $x \in S$ is a boundary point of W if every neighbourhood N_x of x in S contains points both from W and from the complementary set $S \setminus W$. The collection of boundary points for W is called *the boundary* of W, and is denoted by ∂W.

We notice that a boundary point of W is not necessarily contained in W. As an example, a is a boundary point of the half open interval $]a, b]$ in $\mathbb{R}$, and a does not belong to $]a, b]$, while the boundary point b does belong to $]a, b]$.

Usually, the boundary of a set corresponds to what one intuitively imagines. But sometimes life offers surprises. Consider for instance the subset $\mathbb{Q}$ of $\mathbb{R}$, which consists of the rational numbers. Here it is easy to see that all points in $\mathbb{Q}$ are boundary points for $\mathbb{Q}$. But it is even worse. Actually every point in $\mathbb{R}$ is a boundary point of $\mathbb{Q}$. Hence a subset W of a topological space S can consist of nothing but boundary points, and even more surprisingly, it can even happen that $\partial W = S$.

The interior, the exterior, and the boundary of W give a partitioning of S into disjoint sets. We have

$$S = \text{int } W \sqcup \text{int } (S \setminus W) \sqcup \partial W,$$

where $\sqcup$ indicates that the three sets int W, int $(S \setminus W)$ and ∂W are pairwise disjoint.

In concrete situations some of the sets can be empty. We just gave an example for which int $W = $ int $(S \setminus W) = \emptyset$.

1.7 Closed sets

Let S be a topological space.

Definition 1.7.1. A subset A of S is said to be *closed* if the complementary set $S \setminus A$ of A is an open set in S.

In general a subset of S need not be neither open nor closed. On the other hand, it is also possible for a subset to be both open and closed. This actually is the case for any subset in a set equipped with the discrete topology. In other words, one should not make the mistake of thinking that a closed set is just a set which is not open.

Theorem 1.7.2. *A subset A in the topological space S is closed if and only if A contains all its boundary points.*

Proof. If A is a closed set, the complementary set $S \setminus A$ is open and consists of nothing but exterior points of A. But then all boundary points of A have to be contained in A.

Conversely, if all boundary points of A are contained in A, then all points in $S \setminus A$ must be exterior points of A, so that $S \setminus A = \text{int}\,(S \setminus A)$. But then $S \setminus A$ is an open set according to Theorem 1.6.2, and hence A is closed. □

The system of closed sets in a topological space S has the following 'complementary' properties to the system of open sets:

A1 If $\{A_i \mid i \in I\}$ is an arbitrary system of closed sets in S, then the intersection $\bigcap \{A_i \mid i \in I\}$ is also a closed set in S.

A2 If $A_1, \ldots, A_k$ is an arbitrary finite system of closed sets in S, then the union $\bigcup_{i=1}^{k} A_i$ is also a closed set in S.

A3 The empty set $\emptyset$ and the set S itself are closed sets in S.

The properties A1, A2, and A3 follow from the corresponding properties TOP 1, TOP 2, and TOP3 by using the following formulae from set theory:

$$S \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (S \setminus A_i) \quad \text{and} \quad S \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (S \setminus A_i)\,.$$

Under the formation of a complementary set there is complete duality between the concepts of open and closed sets. Indeed, for any subsets U and A in the topological space S it is obvious that:

$$U \quad \text{open} \quad \Longleftrightarrow \quad S \setminus U \quad \text{closed}$$

and

$$A \quad \text{closed} \quad \Longleftrightarrow \quad S \setminus A \quad \text{open}.$$

The main reason for introducing the system of closed sets in a topological space is that it is easier to use in certain situations. In this context the following result is of interest.

Theorem 1.7.3. *A mapping $f : X \to Y$ between topological spaces X and Y is continuous if and only if for every closed set A in Y, the set $f^{-1}(A)$ is closed in X.*

Proof. First, assume that $f : X \to Y$ is continuous, and let A be a closed set in Y. Then $Y \setminus A$ is open in Y, and since f is continuous, the set $X \setminus f^{-1}(A) = f^{-1}(Y \setminus A)$ is therefore open in X. But then $f^{-1}(A)$ is a closed set in X, as was to be proved.

Next assume conversely that $f^{-1}(A)$ is a closed set in X for any closed set A in Y. Let V be any open set in Y. We have to show that $f^{-1}(V)$ is an open set in X. But because $X \setminus f^{-1}(V) = f^{-1}(Y \setminus V)$ this follows immediately, since $Y \setminus V$ is a closed set in Y, and therefore by the assumption that $f^{-1}(Y \setminus V)$ is a closed set in X. □

1.8 The closure of a set

Let W be an arbitrary subset of the topological space S. In Theorem 1.6.2 we have seen that W contains a 'largest' open set, namely int W. We shall now show that correspondingly, W is contained in a 'smallest' closed set $\overline{W}$, which we call the closure of W.

As a preparation we first introduce two types of points in relation to W considered as a subset of S.

Definition 1.8.1. A point $x \in S$ is called a *contact point* of W, provided that every neighbourhood N of x in S contains at least one point from W. A point $x \in S$ is called an *accumulation point* of W, provided that every neighbourhood N of x in S contains at least one point from W different from x.

Clearly interior points and boundary points of W are contact points of W. Conversely, it is also immediate from the definitions, that a contact point of W is either an interior point or a boundary point of W. In other words, *'contact point' is a common name for interior point and boundary point,* and therefore we have in fact not really introduced a new concept by this definition. In many cases, however, it is convenient to have the common notion.

A contact point of W which does not belong to W has to be an accumulation point. An *isolated point* in W is a contact point of W which is not an accumulation point of W. An isolated point in W is in other words a point $x \in W$, which has a neighbourhood N_x in S, in which x is the only point from W.

Since interior points of a subset naturally belong to the subset, we can now give Theorem 1.7.2 a more convenient formulation.

Theorem 1.8.2. *A subset A in the topological space S is closed if and only if A contains all its contact points (accumulation points).*

It is now clear what we have to do to in order to 'close' a subset.

Definition 1.8.3. The collection of all contact points of the subset W in the topological space S is called *the closure* of W and is denoted by $\overline{W}$.

It appears that $\overline{W}$ is obtained by adding to W those contact points for W which do not already belong to W.

From the description of contact points as either interior points or boundary points for W it follows immediately that

$$\overline{W} = \text{int } W \cup \partial W.$$

Hence one gets $\overline{W}$ by adding to W the boundary points of W which do not already belong to W.

Theorem 1.8.4. *The closure of a subset W in the topological space S has the following properties:*

(i) *$\overline{W}$ is a closed set.*

(ii) *If A is a closed set in S so that $W \subseteq A$, then $\overline{W} \subseteq A$. ($\overline{W}$ is the 'smallest' closed set in S which contains W.)*

(iii) *If A is a closed set in S so that $W \subseteq A \subseteq \overline{W}$, then $A = \overline{W}$.*

(iv) *$\overline{W}$ is the intersection of all closed sets A in S for which $W \subseteq A$.*

Proof. Proof of (i). By using

$$\overline{W} = \text{int } W \cup \partial W = S \setminus \text{int } (S \setminus W),$$

it follows immediately from Theorem 1.6.2 (i) that $\overline{W}$ is closed.

Proof of (ii). Let A be a closed set in S such that $W \subseteq A$. Let $x \in \overline{W}$ be an arbitrary point in $\overline{W}$. Since x is a contact point of W, any open neighbourhood of x in S will contain at least one point from W, thus also from A since $W \subseteq A$. But then x is a contact point for A, and since A is closed it follows that $x \in A$ by Theorem 1.8.2. This proves that $\overline{W} \subseteq A$.

Proof of (iii). Follows immediately from (ii).

Proof of (iv). The intersection of all closed sets A in S such that $W \subseteq A$ is itself a closed set $\tilde{A}$ in S such that $W \subseteq \tilde{A}$. According to (i), $\overline{W}$ is itself one of the sets over which the intersection is taken, hence $W \subseteq \tilde{A} \subseteq \overline{W}$. But then according to (iii) we have that $\tilde{A} = \overline{W}$. □

Theorem 1.8.4 has the following immediate corollary.

Corollary 1.8.5. *A subset A in a topological space S is closed if and only if $A = \overline{A}$.*

We have the following simple results regarding the effect of taking the closure of the union, respectively the intersection, of two sets.

Theorem 1.8.6. *For subsets W_1 and W_2 in the topological space S the following two statements hold:*

$$\text{(a)} \quad \overline{W_1 \cup W_2} = \overline{W_1} \cup \overline{W_2} \ .$$
$$\text{(b)} \quad \overline{W_1 \cap W_2} \subseteq \overline{W_1} \cap \overline{W_2} \ .$$

Proof. Since $W_1 \cup W_2 \subseteq \overline{W_1} \cup \overline{W_2}$ and since $\overline{W_1} \cup \overline{W_2}$ is a closed set in S, it follows from Theorem 1.8.4 (ii) that,

$$\overline{W_1 \cup W_2} \subseteq \overline{W_1} \cup \overline{W_2}.$$

Since a contact point of either W_1 or W_2 is clearly a contact point of $W_1 \cup W_2$, we also get the opposite inclusion,

$$\overline{W_1} \cup \overline{W_2} \subseteq \overline{W_1 \cup W_2}.$$

This proves part (a).

A contact point for $W_1 \cap W_2$ is clearly a contact point of both W_1 and W_2, which proves part (b). □

Remark 1.8.7. On the real axis $\mathbb{R}$ consider the subsets

$$W_1 = \{x \in \mathbb{R} \mid x < 0\} \quad \text{and} \quad W_2 = \{x \in \mathbb{R} \mid x > 0\}.$$

Then $\overline{W_1 \cap W_2} = W_1 \cap W_2 = \emptyset$, whereas $\overline{W_1} \cap \overline{W_2} = \{0\}$. Therefore $\overline{W_1 \cap W_2}$ is, in general, a proper subset of $\overline{W_1} \cap \overline{W_2}$ in Theorem 1.8.6 (b).

In the theory of singularities, as well as in the qualitative theory of dynamical systems, one is interested in *generic* ('typical') properties of mappings or dynamical systems. Informally speaking, a property is said to be generic if it is valid for a dense subset of the objects under consideration (mappings or dynamical systems) when considered as points in an appropriate topological space.

Definition 1.8.8. A subset W in a topological space S is said to be *dense* in S if $\overline{W} = S$.

The concept of density can be given many equivalent formulations. We list some of them in the following theorem, the proof of which is left to the reader.

Theorem 1.8.9. *For a subset W in a topological space S the following statements are equivalent:*

(i) *W is dense in S.*

(ii) *Every point in S either belongs to W or is an accumulation point of W.*

(iii) *Every non-empty, open subset in S contains at least one point from W.*

(iv) $\text{int}\ (S \setminus W) = \emptyset$.

It is also of considerable interest to know whether a property of a mapping, or a dynamical system, is *stable* (robust under small perturbations). From a topological point of view, a property is stable if the objects with the desired property form an open set, when considered as points in an appropriate topological space.

Example 1.8.10. Let $\mathbb{Z}$ denote the integers and consider the subset $W = \mathbb{R}\setminus\mathbb{Z}$ on the real axis $\mathbb{R}$. Then W is an open and dense subset of $\mathbb{R}$. So for a real number it is both a generic and a stable property not to be an integer. The subset $\mathbb{I}$ of irrational numbers in $\mathbb{R}$ is dense but not open. So for a real number it is a generic property, but not a stable property, to be an irrational number.

1.9 Limit points. Hausdorff spaces

The concept of a limit point for a mapping or a sequence, and the corresponding concept of convergence, can easily be introduced in the setting of topological spaces. We introduce these concepts simultaneously in metric spaces and in topological spaces, so that one can easily see the relation to the well-known situations from the Euclidean spaces.

Definition 1.9.1 (metric). Let $f : X \to Y$ be a mapping between metric spaces (X, d_X) and (Y, d_Y), and let $x_0 \in X$ and $y_0 \in Y$. We say that

$$f(x)\ \textit{approaches}\ y_0\ \text{for}\ x\ \textit{approaching}\ x_0\ ,$$

or that

$$f(x)\ \text{has the}\ \textit{limit point}\ y_0\ \text{for}\ x\ \textit{approaching}\ x_0\ ,$$

or that

$$f(x)\ \textit{converges to}\ y_0\ \text{for}\ x\ \textit{converging to}\ x_0\ ,$$

if:

$$\forall \varepsilon > 0 \ \ \exists \delta > 0 \ \ \forall x \in X \setminus \{x_0\} : \ d_X(x, x_0) < \delta \ \Rightarrow \ d_Y(f(x), y_0) < \varepsilon \ .$$

Definition 1.9.1 (topological). Let $f : X \to Y$ be a mapping between topological spaces X and Y, and let $x_0 \in X$ and $y_0 \in Y$. We say that

$$f(x) \ \textit{approaches} \ y_0 \ \text{for} \ x \ \textit{approaching} \ x_0 \ ,$$

or one of the other similar terminologies presented in the metric case, if:

For any neighbourhood M of y_0 in Y there exists a neighbourhood N of x_0 in X such that

$$f(N \setminus \{x_0\}) \subseteq M.$$

For sequences of points we have similar concepts.

Definition 1.9.2 (metric). Let (x_n), or more explicitly $x_1, x_2, \ldots, x_n, \ldots$, be a sequence of points in the metric space (S, d), and let $y_0 \in S$. We say that

$$x_n \ \textit{approaches} \ y_0 \ \text{for} \ n \ \textit{going to} \ \infty \ ,$$

or that

$$x_n \ \text{has the} \ \textit{limit point} \ y_0 \ \text{for} \ n \ \textit{going to} \ \infty \ ,$$

or that

$$x_n \ \textit{converges to} \ y_0 \ \text{for} \ n \ \textit{going to} \ \infty \ ,$$

if :

$$\forall \varepsilon > 0 \ \ \exists n_0 \in \mathbb{N} \ \ \forall n \in \mathbb{N} : \ n \geq n_0 \quad \Rightarrow \quad d(x_n, y_0) < \varepsilon \ .$$

Definition 1.9.2 (topological). Let (x_n), or more explicitly x_1, $x_2, \ldots, x_n$, $\ldots$, be a sequence of points in the topological space S, and let $y_0 \in S$. We say that

$$x_n \ \textit{approaches} \ y_0 \ \text{for} \ n \ \textit{going to} \ \infty \ ,$$

or one of the other similar terminologies presented in the metric case, if:

For every neighbourhood M of y_0 in S there exists an $n_0 \in \mathbb{N}$ such that

$$x_n \in M \quad \text{for} \quad n \geq n_0 \ .$$

Often, we just say that the sequence (x_n) in the topological space S is *convergent* with the *limit point* y_0, or that the sequence (x_n) *converges* to y_0.

In the two situations described in Definition 1.9.1, respectively Definition 1.9.2, we use the short notations:

$$f(x) \to y_0 \quad \text{for} \quad x \to x_0 \ , \quad \text{or} \quad \lim_{x \to x_0} f(x) = y_0 \ ,$$

respectively

$$x_n \to y_0 \quad \text{for} \quad n \to \infty\,, \quad \text{or} \quad \lim_{n\to\infty} x_n = y_0\,.$$

The definitions of continuity and convergence are closely related, and in the setting of metric spaces continuity of a mapping at a point can in fact be described by convergence of sequences.

Theorem 1.9.3. *A mapping $f : X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is continuous at a point $x_0 \in X$ if and only if for every convergent sequence (x_n) of points in X with limit point x_0, the sequence $(f(x_n))$ of points in Y is convergent with limit point $f(x_0)$.*

Proof. First, assume that $f : X \to Y$ is continuous at $x_0 \in X$. Let (x_n) be a convergent sequence of points in X with limit point x_0. We have to show that $(f(x_n))$ is a convergent sequence of points in Y with limit point $f(x_0)$. For this, let an arbitrary $\varepsilon > 0$ be given. Since f is continuous at x_0, we can choose $\delta > 0$, such that for $x \in X$ with $d_X(x, x_0) < \delta$ we have $d_Y(f(x), f(x_0)) < \varepsilon$. Because the sequence (x_n) is convergent with limit point x_0, for $\delta > 0$ we can choose an $n_0 \in \mathbb{N}$ such that $d_X(x_n, x_0) < \delta$ for $n \geq n_0$. But then $d_Y(f(x_n), f(x_0)) < \varepsilon$, for $n \geq n_0$. This shows that $(f(x_n))$ converges to $f(x_0)$.

Now assume, conversely, that for every convergent sequence (x_n) of points in X with limit point x_0, we know that the sequence $(f(x_n))$ of points in Y converges towards $f(x_0)$. We have to show that f is continuous at x_0. Then:

$$\exists \varepsilon_0 > 0 \;\; \forall \delta > 0 \;\; \exists x \in X : \;\; (d_X(x, x_0) < \delta) \wedge (d_Y(f(x), f(x_0)) \geq \varepsilon_0).$$

Consider such an $\varepsilon_0 > 0$, and choose for every $n \in \mathbb{N}$ a point $x_n \in X$ such that

$$(d_X(x_n, x_0) < 1/n) \quad \text{and} \quad (d_Y(f(x_n), f(x_0)) \geq \varepsilon_0).$$

Now, it is clear that (x_n) is a convergent sequence of points in X with limit point x_0, even though $(f(x_n))$ does not converge to $f(x_0)$. This yields a contradiction, which proves that f is continuous at x_0. ☐

In a general topological space, a convergent sequence of points may have several limit points. Usually we want to avoid this, and therefore we need to require more of the topological space. In that connection a suitable separation axiom for the points in S was formulated in the book "Grundzüge der Mengenlehre" of 1914 by the German mathematician Felix Hausdorff (1868–1942), who is one of the pioneers in general topology.

Definition 1.9.4. A topological space S is called a *Hausdorff space*, if for every pair of distinct points x and y in S there exists a related pair of disjoint open neighbourhoods U of x and V of y.

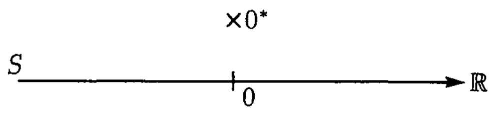


Figure 1.6

Theorem 1.9.5. *A metric space (S, d) is a Hausdorff space.*

Proof. Let x and y be an arbitrary pair of distinct points in S. Then

$$r = d(x, y) > 0.$$

The triangle inequality now implies that the open sets $U = B_{r/2}(x)$ and $V = B_{r/2}(y)$ are disjoint open neighbourhoods of x and y, as required. □

Theorem 1.9.6. *A sequence (x_n) of points in a Hausdorff space S has at most one limit point.*

Proof. The proof is by contradiction. So assume that the sequence (x_n) of points in the Hausdorff space S admits two distinct limit points x and y. Corresponding to x and y there is a pair of disjoint, open neighbourhoods U of x and V of y. Since (x_n) is convergent with limit point x, respectively y, we can choose $n_1 \in \mathbb{N}$, respectively $n_2 \in \mathbb{N}$, such that $x_n \in U$ for $n \geq n_1$, respectively $x_n \in V$ for $n \geq n_2$. For $n \geq \max(n_1, n_2)$ this yields a contradiction, since $x_n \in U \cap V$ and $U \cap V = \emptyset$. □

We finish this section with an example illustrating the idea of the Hausdorff axiom.

Example 1.9.7 (Line with an extra origin). Consider the set S defined by adding an extra origin 0^* to the real axis $\mathbb{R}$.

We equip S with a topology, in which the open sets are the usual open sets in $\mathbb{R}$ together with those subsets U of S that contain 0^* and have the property that $(U \setminus \{0^*\}) \cup \{0\}$ is an ordinary open set in $\mathbb{R}$. It is easy to prove that this actually defines a topology on S.

The set S equipped with the above topology is not a Hausdorff space. Indeed, it is clear that we cannot separate the points 0 and 0^* in S with a pair of disjoint, open subsets. Furthermore, it is clear that the sequence $(1/n)$ admits both 0 and 0^* as limit points.

The topological space S is, in other words, not a Hausdorff space. It has, however, a weaker separation property, which is also interesting, namely the following property.

Separation Axiom T_1. *For each pair of distinct points x and y in S there exists an open neighbourhood U of x which does not contain y, and an open neighbourhood V of y which does not contain x.*

A topological space satisfying the separation axiom T_1 is called a *T_1-space.* Correspondingly, the separation axiom behind the notion of a Hausdorff space is called T_2, and the space itself a *T_2-space.* The term T-space was introduced by the Russian mathematician Pavel S. Aleksandrov (1896–1982) and the German mathematician Heinz Hopf (1894–1971) in their book "Topologie I" of 1935. (T stands for the German word 'Trennung', which means separation.)

1.10 Compact sets

We start out by stating two classical theorems concerning closed and bounded intervals in $\mathbb{R}$.

Theorem 1.10.1 (Bolzano-Weierstrass). *Every infinite subset A of points in a closed and bounded interval $[a, b]$ of $\mathbb{R}$ has at least one accumulation point in $[a, b]$.*

Theorem 1.10.2 (Heine-Borel). *Every covering of a closed and bounded interval $[a, b]$ of $\mathbb{R}$ by a system of open intervals $\{U_i \mid i \in I\}$ contains a finite subcovering.*

Theorem 1.10.2 asserts: If $\{U_i \mid i \in I\}$ is any system of open intervals in $\mathbb{R}$ such that $[a, b] \subseteq \cup_{i \in I} U_i$, then we can extract finitely many intervals from the system, say $U_{i_1}, U_{i_2}, \ldots, U_{i_n}$, such that $[a, b] \subseteq \cup_{k=1}^{n} U_{i_k}$.

Theorem 1.10.1 was anticipated by the Bohemian priest and mathematician Bernhard Bolzano (1781–1848) in 1817 and fully developed by the German mathematician Karl Weierstrass (1815–1897) in the 1860s.

The covering property in Theorem 1.10.2 was used by the German mathematician Eduard Heine (1821–1881) in 1872 in a study of uniform continuity and fully recognized as an important property by the French mathematician Emile Borel (1871–1956) in 1895 for countable open coverings, and by his pupil, the French mathematician Henri Lebesque (1875–1941), in 1904 for general open coverings.

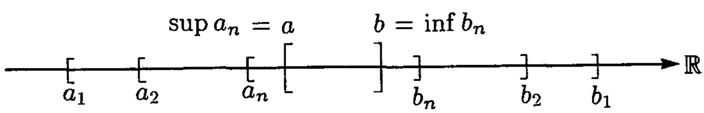


Figure 1.7

Both of the above theorems are based on the following basic property of the real numbers, which is equivalent to the existence of supremum (and hence also to the existence of infimum).

Property 1.10.3 (The principle of nested intervals).
Let $[a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots \supseteq [a_n, b_n] \supseteq \ldots$ *be a decreasing sequence of closed and bounded intervals of* $\mathbb{R}$. *Then:*

(i) *The intersection* $\bigcap_{n=1}^{\infty}[a_n, b_n]$ *is a closed and bounded interval of* $\mathbb{R}$.

(ii) *If the length of the interval* $|b_n - a_n|$ *approches* 0 *for increasing* n *then the intersection* $\bigcap_{n=1}^{\infty}[a_n, b_n]$ *contains exactly one real number.*

A nested interval sequence $[a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots \supseteq [a_n, b_n]$ in which $|b_n - a_n| \to 0$ for $n \to \infty$, we call, for short, an *interval trap*.

Since we have taken the postulate of existence of supremum as the basic characteristic property of the real numbers, we prove for completeness that the above property of the real numbers is a consequence of this postulate.

Proof. First we prove (i). Since the sequence (a_n) is bounded above, e.g. by b_1, the number $a = \sup a_n$ exists. Analogously, the number $b = \inf b_n$ exists, since the sequence (b_n) is bounded below, e.g. by a_1. Because every b_n is an upper bound for (a_n), we have that $a \leq b_n$ for all $n \in \mathbb{N}$. But then a is a lower bound for (b_n), from which it follows that $a \leq b$; cf. Figure 1.7. So what is left is to show that $[a, b] = \cap_{n=1}^{\infty}[a_n, b_n]$. The inclusion $\subseteq$ is immediate, since $a_n \leq a \leq b \leq b_n$ for all $n \in \mathbb{N}$. The inclusion $\supseteq$ follows by observing that a number c satisfying $a_n \leq c \leq b_n$ for all $n \in \mathbb{N}$ is an upper bound for (a_n) and a lower bound for (b_n) and therefore has to satisfy $a \leq c \leq b$.

Next we prove (ii). Since $[a, b] \subseteq [a_n, b_n]$ for all $n \in \mathbb{N}$, it is clear that $a = b$ when $|b_n - a_n| \to 0$ for $n \to \infty$. In this situation we therefore have

$$\bigcap_{n=1}^{\infty}[a_n, b_n] = \{a\} .$$

This completes the proof. □

Conversely, it is also easy to deduce the existence of supremum from the principle of nested intervals. We leave this to the reader.

Proof of Theorem 1.10.1. Let A be an infinite subset of points in the closed and bounded interval $[a, b]$ in $\mathbb{R}$. By successively partitioning the interval at the middle we can construct an interval trap

$$[a, b] = [a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots \supseteq [a_n, b_n] \supseteq \dots ,$$

in which every interval $[a_n, b_n]$ contains infinitely many points from A. The interval trap determines a real number c, such that

$$\{c\} = \bigcap_{n=1}^{\infty} [a_n, b_n] .$$

Any neighbourhood M of c contains all the intervals $[a_n, b_n]$ from a certain step n_0, thus infinitely many points from A. In particular, M contains at least one point from A distinct from c. Hence the point $c \in [a, b]$ is an accumulation point of A. □

Proof of Theorem 1.10.2. The proof is by contradiction. So assume that there exists a system of open intervals $\{U_i \mid i \in I\}$ which covers the closed and bounded interval $[a, b]$ in $\mathbb{R}$, but which does not contain a finite subcovering of $[a, b]$. By successively partitioning the interval at the middle we can construct an interval trap

$$[a, b] = [a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots \supseteq [a_n, b_n] \supseteq \dots ,$$

in which none of the intervals $[a_n, b_n]$ can be covered by finitely many intervals from the system $\{U_i \mid i \in I\}$. The interval trap determines a real number c, such that

$$\{c\} = \bigcap_{n=1}^{\infty} [a_n, b_n] .$$

Since $c \in [a, b]$, there is an interval U_{i_0} from the system $\{U_i \mid i \in I\}$, such that $c \in U_{i_0}$. Since U_{i_0} is an open interval, $[a_n, b_n] \subseteq U_{i_0}$ from a certain step n_0. This contradicts the construction of the intervals $[a_n, b_n]$. Our assumption led to a contradiction, and consequently from any covering of $[a, b]$ with open intervals we can extract a finite subcovering. □

We shall now make a brief detour from the main theme of this section and remind the reader that a set is said to be *countable*, if the elements in the set can be indexed by the set of natural numbers $\mathbb{N}$. At the end of section 1.5.3 we indicated that there are sets which are not countable. Using interval traps, we can prove the following theorem in connection with this.

Theorem 1.10.4. *The set of real numbers $\mathbb{R}$ is not countable.*

Proof. The proof is by contradiction. Assume therefore that $\mathbb{R}$ is a countable set, and that $r_1, r_2, \ldots, r_n, \ldots$ is a list of all the real numbers. First, choose a closed and bounded interval $[a_1, b_1]$ in $\mathbb{R}$ which does not contain r_1. Then partition the interval $[a_1, b_1]$ into three equally long subintervals. At least one of these subintervals, which we choose and denote by $[a_2, b_2]$, does not contain r_2. By successively partitioning one of the intervals into three subintervals as indicated, we can construct an interval trap

$$[a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots \supseteq [a_n, b_n] \supseteq \cdots ,$$

in which the interval $[a_n, b_n]$ does not contain the real number r_n. The interval trap determines a real number c, such that

$$\{c\} = \bigcap_{n=1}^{\infty} [a_n, b_n] .$$

Clearly $c \neq r_n$ for all $n \in \mathbb{N}$. This is in contradiction with the assumption that $r_1, r_2, \ldots, r_n, \ldots$ is a list of all real numbers. We conclude that $\mathbb{R}$ is not countable. □

Using either Theorem 1.10.1 or Theorem 1.10.2 we can easily prove the following fundamental result.

Theorem 1.10.5. *Let $[a, b]$ be a closed and bounded interval of $\mathbb{R}$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is bounded. In other words, there exists a constant k, such that $|f(x)| \leq k$ for all $x \in [a, b]$.*

Proof by using Theorem 1.10.1. The proof is by contradiction. Thus we assume that f is not bounded. Then we can find $x_1 \in [a, b]$ with $|f(x_1)| > 1$, because f would otherwise be bounded by 1. Now, we can find $x_2 \in [a, b]$ with $x_2 \neq x_1$ such that $|f(x_2)| > 2$, because f would otherwise be bounded by $\max\{2, |f(x_1)|\}$. Recursively, we can produce a sequence $x_1, x_2, \ldots, x_n, \ldots$ of mutually distinct points in $[a, b]$ with $|f(x_n)| > n$ for all $n \in \mathbb{N}$. According to Theorem 1.10.1 the infinite set $\{x_n\}$ has an accumulation point x_0 in $[a, b]$. (Note that it is not necessarily true that $x_n \to x_0$ as $n \to \infty$.) Since f is continuous at x_0, there exists a $\delta > 0$, such that $|f(x)| \leq |f(x_0)| + 1$ for $|x - x_0| < \delta$. Now, choose an integer m such that $m > |f(x_0)| + 1$. Since x_0 is an accumulation point of $\{x_n\}$ there exists a point x_n with index $n > m$ such that $|x_n - x_0| < \delta$. But this yields a contradiction, since $n > m$ implies that $|f(x_n)| > n > m$, and $|x_n - x_0| < \delta$ implies that $|f(x_n)| \leq |f(x_0)| + 1 < m$. Hence our assumption led to a contradiction, and thus f must be bounded. □

Proof by using Theorem 1.10.2. Since f is continuous we can, for every $x \in [a, b]$, choose a $\delta_x > 0$ such that $|f(y)| \leq |f(x)| + 1$ for $|y - x| < \delta_x$. Let U_x be the open interval

$$U_x =]x - \delta_x, x + \delta_x[\ .$$

The system of open intervals $\{U_x \mid x \in [a, b]\}$ covers $[a, b]$. By Theorem 1.10.2 finitely many of these intervals, say $U_{x_1}, U_{x_2}, \ldots, U_{x_n}$, already cover $[a, b]$. Now, put

$$k = \max\{|f(x_1)|, |f(x_2)|, \ldots, |f(x_n)|\} + 1\ .$$

Since $U_{x_1}, U_{x_2}, \ldots, U_{x_n}$ cover $[a, b]$, there exists for every $x \in [a, b]$ at least one x_i, such that $x \in U_{x_i}$, or equivalently $|x - x_i| < \delta_{x_i}$. But then we have immediately that

$$|f(x)| \leq |f(x_i)| + 1 \leq k.$$

Thus f is bounded. □

The results stated in Theorem 1.10.1 and Theorem 1.10.2 are essential, and they motivate the introduction of certain concepts that have proved fundamental in mathematical analysis.

With the theorem of Bolzano and Weierstrass (Theorem 1.10.1) as inspiration we introduce the concept of 'sequentially compact subsets' of a topological space.

Definition 1.10.6. Let S be a topological space. A subset K in S is called *sequentially compact* if every infinite subset A of points in K has an accumulation point in K.

We notice that, according to Theorem 1.10.1, any closed and bounded interval $[a, b]$ is a sequentially compact subset of $\mathbb{R}$.

The terminology 'sequentially compact' is perhaps best justified by the following theorem, the proof of which is left to the reader.

Theorem 1.10.7. *A subset K in a metric space S is sequentially compact if and only if every sequence (x_n) of points in K contains a convergent subsequence (x_{n_k}) with limit point x_0 in K.*

Inspired by the Theorem of Heine and Borel (Theorem 1.10.2), we introduce the concept of a 'compact subset' in a topological space. In this book, we shall mainly be concerned with the concept of compactness.

First we need a definition.

Definition 1.10.8. Let S be a topological space, and let K be a subset of S.

By an *open covering* of K we understand a system of open sets $\{U_i \mid i \in I\}$ in S such that $K \subseteq \bigcup_{i \in I} U_i$.

By a *subcovering* of an open covering $\{U_i \mid i \in I\}$ of K we understand a subfamily of the open sets $\{U_i \mid i \in I\}$ which itself covers K.

An important issue is whether finite subcoverings can be extracted from open coverings. This leads to the notion of compactness.

Definition 1.10.9. Let S be a topological space. A subset K in S is called *compact* if every open covering $\{U_i \mid i \in I\}$ of K contains a finite subcovering.

We note that, according to Theorem 1.10.2, any closed and bounded interval $[a, b]$ is a compact subset of $\mathbb{R}$.

1.11 Compact sets in Euclidean spaces

The main result of this section states that the compact subsets of $\mathbb{R}^n$ are precisely the subsets which are both closed and bounded. This way we obtain a perfect generalization of Theorem 1.10.2.

We approach this characterization of the compact subsets of $\mathbb{R}^n$ via two theorems which remain valid in more general topological spaces.

Theorem 1.11.1. *In a Hausdorff space S every compact subset K is closed.*

Proof. Let K be a compact subset in the Hausdorff space S, and consider the complement $S \setminus K$. We have to show that $S \setminus K$ is an open set in S. If $S \setminus K = \emptyset$ there is nothing to prove. So we assume that $S \setminus K \neq \emptyset$ and consider an arbitrary point $x \in S \setminus K$. Since S is a Hausdorff space, we can, for every point $y \in K$, choose a pair of disjoint open sets U_y and V_y in S, such that $x \in U_y$ and $y \in V_y$. The system of open sets $\{V_y \mid y \in K\}$ is an open covering of K. From this covering we can extract a finite subcovering, say $V_{y_1}, V_{y_2}, \dots, V_{y_n}$, of K, since K is compact. Since $U_{y_i} \cap V_{y_i} = \emptyset$ for every $i = 1, \dots, n$, the set $U = U_{y_1} \cap \dots \cap U_{y_n}$ is an open neighbourhood of x in S, such that $U \subseteq S \setminus K$. According to Lemma 1.5.3, this shows that $S \setminus K$ is an open set in S, and hence that K is a closed set in S. □

Definition 1.11.2. A subset K in a metric space (S, d) is called *bounded*, if it is completely contained in an open ball in S.

Theorem 1.11.3. *In a metric space (S, d) every compact subset K is bounded.*

Proof. Let K be a compact subset in the metric space (S, d), and let $x_0 \in S$ be a fixed point in S. Consider the system of open balls $\{B_n(x_0) \mid n \in \mathbb{N}\}$ in S. Since every point in S is contained in these balls for n sufficiently large, it is clear that the system is an open covering of K. Since K is compact, finitely many of these balls will cover K, and hence obviously K is contained in that ball, among these finitely many balls, which has the largest radius. This proves that K is bounded. □

Theorems 1.11.1 and 1.11.3 show that in an arbitrary metric space every compact subset is closed and bounded. If, in particular, $\mathbb{R}^n$ is considered with the usual topology induced by the Euclidean metric, then the converse is also true. Thereby we obtain the following main result.

Theorem 1.11.4 (Heine-Borel). *For a subset K in $\mathbb{R}^n$ the following statements are equivalent:*

(1) *K is compact: Every open covering of K contains a finite subcovering.*

(2) *K is closed and bounded.*

Theorem 1.11.4 is like Theorem 1.10.2 due to Heine and Borel, with a refinement by Lebesque. The power of the theorem lies mainly in the abstract ideas developed, and as earlier mentioned, the theorem crystallized in its final form only shortly after 1900.

Proof of Theorem 1.11.4. The implication (1)⇒(2) has already been proved; cf. Theorems 1.11.1 and 1.11.3.

It remains to prove (2)⇒(1). Assume therefore that K is a closed and bounded subset in $\mathbb{R}^n$. We shall prove that K is compact. The proof is by contradiction, and hence we assume that there exists an open covering $\{U_i \mid i \in I\}$ of K from which we cannot extract a finite subcovering. As we shall see this leads to a contradiction.

Since K is bounded, we may choose an n-dimensional box

$$C^1 = [a_1^1, b_1^1] \times [a_2^1, b_2^1] \times \cdots \times [a_n^1, b_n^1]$$

such that $K \subseteq C^1$. We now partition each of the intervals $[a_i^1, b_i^1]$ at the middle, which yields 2^n new small boxes. Among these small boxes, there must be at least one, which we choose and denote by C^2, for which $C^2 \cap K$ cannot be covered by finitely many open sets from the system $\{U_i \mid i \in I\}$; cf. Figure 1.8.

By successively bisecting the intervals, we can construct a decreasing nested sequence of n-dimensional boxes

$$C^1 \supseteq C^2 \supseteq \cdots \supseteq C^k \supseteq \dots ,$$

in which it holds for every $k \in \mathbb{N}$ that $C^k \cap K$ cannot be covered by finitely many open sets from the system $\{U_i \mid i \in I\}$.

If we write the box C^k in the form

$$C^k = [a_1^k, b_1^k] \times [a_2^k, b_2^k] \times \cdots \times [a_n^k, b_n^k] ,$$

it is clear that for every $i = 1, \dots n$ we get an interval trap

$$[a_i^1, b_i^1] \supseteq [a_i^2, b_i^2] \supseteq \cdots \supseteq [a_i^k, b_i^k] \supseteq \dots .$$

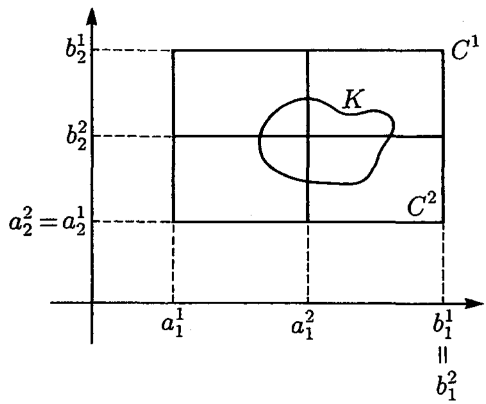


Figure 1.8

This interval trap determines a real number c_i.

Now, consider the point $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$. It is clear that

$$\bigcap_{k=1}^{\infty} C^k = \{c\}.$$

Every open neighbourhood of c in $\mathbb{R}^n$ therefore contains all the boxes C^k from a certain step k_0, thus in particular it contains infinitely many points from K. Indeed, if this was not the case, $C^k \cap K$ could have been covered by finitely many open sets from the system $\{U_i \mid i \in I\}$. Hence c is an accumulation point of K. Since K is closed, it follows that $c \in K$ by Theorem 1.8.2.

Since $c \in K$ and the system $\{U_i \mid i \in I\}$ covers K, there must be an $i_0 \in I$ such that $c \in U_{i_0}$. From a certain step k_0, all the boxes C^k will be completely contained in the open neighbourhood U_{i_0} of c. But this contradicts that $C^k \cap K$ could not be covered by finitely many open sets from the system $\{U_i \mid i \in I\}$. This completes the proof. □

Example 1.11.5. Let S be an arbitrary infinite set, and equip S with the discrete metric. Then every subset of S is both open and closed. Furthermore, it is bounded, because it is contained in any open ball of say radius 2. In particular, we note that every subset of S is both closed and bounded. However, an infinite subset A in S is not compact. For example $\{U_a = \{a\} \mid a \in A\}$ is an open covering of A which does not contain a finite subcovering. Therefore, in general, a closed and bounded subset of a metric space is not compact. This property is a particular property of $\mathbb{R}^n$.

1.12 Infinite subsets of compact sets

In this section we shall study the relationship between the concepts compact and sequentially compact. The main result is, that in metric spaces these concepts coincide. First, however, we show that in complete generality compact implies sequentially compact.

Theorem 1.12.1. *Let K be a compact subset in the topological space S. Then every infinite subset A of K has at least one accumulation point in K. In other words: A compact subset K in the topological space S is sequentially compact.*

Proof. The proof is by contradiction. Assume that A is an infinite subset of K without accumulation points in K. Then we can choose an open neighbourhood U_x around every point $x \in K$, such that U_x contains at most one point from A; namely the point x itself if $x \in A$. The open covering $\{U_x \mid x \in K\}$ does not contain a finite subcovering, since A is infinite. This contradicts the hypothesis that K is compact. Hence every infinite subset A in K has at least one accumulation point in K. □

In a metric space the converse is also true, i.e. a sequentially compact subset is compact. This yields the following main theorem.

Theorem 1.12.2. *For a subset K in a metric space (S, d) the following statements are equivalent:*

(1) *K is compact: Every open covering of K contains a finite subcovering.*

(2) *K is sequentially compact: Every infinite subset A in K has at least one accumulation point in K.*

Proof. The implication (1)⇒(2) has already been proved; cf. Theorem 1.12.1

It remains to prove (2)⇒(1). Let therefore $\{U_i \mid i \in I\}$ be an arbitrary open covering of K. We shall prove that we can extract a finite subcovering from this covering. The case $K = \emptyset$ is trivial, so assume that $K \neq \emptyset$.

Assertion 1.12.3. *There exists a positive real number $r \in \mathbb{R}^+$, such that for every point $x \in X$ the ball $B_r(x)$ is contained in one of the sets U_i.*

Proof of Assertion 1.12.3. The proof is by contradiction. Assume therefore that for every $r \in \mathbb{R}^+$ there exists a point $x \in K$ such that $B_r(x)$ is not contained in any of the sets U_i. In particular, then, for every $n \in \mathbb{N}$, we can choose $x_n \in K$, such that $B_{1/n}(x_n)$ is not contained in any of the sets U_i. The set $\{x_n\}$ is an infinite set in K, since from a certain step n_0 any ball $B_{1/n}(x')$ around a fixed point $x' \in K$ is contained in one of the open sets U_i. According to (2), the infinite set $\{x_n\}$ in K has an accumulation point $x_0 \in K$. Since $\{U_i \mid i \in I\}$ covers K, we can first choose $i_0 \in I$ such that $x_0 \in U_{i_0}$, and the set U_{i_0} being open in S, we can next choose $r_0 \in \mathbb{R}^+$ such that $B_{r_0}(x_0) \subseteq U_{i_0}$. Since x_0 is an accumulation point of $\{x_n\}$, the set $\{n \in \mathbb{N} \mid x_n \in B_{r_0/2}(x_0)\}$ is an infinite set of natural numbers, because otherwise we could easily construct an open ball around x_0 not containing any point $x_n \neq x_0$. So we can choose $m \in \mathbb{N}$, such that $x_m \in B_{r_0/2}(x_0)$ and $1/m < r_0/2$. But then, by the triangle inequality,

$$B_{1/m}(x_m) \subseteq B_{r_0}(x_0) \subseteq U_{i_0}.$$

This contradicts the assumption that $B_{1/m}(x_m)$ is not contained in any of the sets U_i, and hence proves Assertion 1.12.3. □

Now, choose a positive real number $r \in \mathbb{R}^+$ in accordance with Assertion 1.12.3, and let $y_1 \in K$ be an arbitrary point in K. If $K \subseteq B_r(y_1)$, then K is covered by one of the sets from $\{U_i \mid i \in I\}$, and we are finished. Otherwise, we choose a point $y_2 \in K \setminus B_r(y_1)$. If $K \subseteq B_r(y_1) \cup B_r(y_2)$, then K is covered by two of the sets from $\{U_i \mid i \in I\}$, and again we are finished. Otherwise, we choose a point $y_3 \in K \setminus (B_r(y_1) \cup B_r(y_2))$, etc ...

Assertion 1.12.4. *The process described above is finite.*

Proof Assertion 1.12.4. If the process was not finite, we would get a sequence (y_n) in K, with $d(y_n, y_m) \geq r$ for $n \neq m$. Such an infinite set $\{y_n\}$ has no accumulation point in K, which contradicts (2). This proves Assertion 1.12.4. □

According to Assertion 1.12.4, we can choose finitely many points $y_1, \ldots, y_p$ in K such that

$$K \subseteq B_r(y_1) \cup \cdots \cup B_r(y_p).$$

But then K is covered by finitely many of the sets from the system $\{U_i \mid i \in I\}$, and the proof of Theorem 1.12.2 is complete. □

Until now we have considered compact subsets only of a topological space. If the whole space itself is compact, then we call it a *compact topological space.* In such a space the following trivial but very useful statement holds.

Theorem 1.12.5. *Let S be a compact topological space. Then every closed subset K in S is compact.*

Proof. Let $\{U_i \mid i \in I\}$ be an arbitrary open covering of K. Since K is closed, $\{U_i \mid i \in I\} \cup \{S \setminus K\}$ is an open covering of S. By the compactness of S this covering contains a finite subcovering of S. Such a finite covering of S must comprise finitely many sets from $\{U_i \mid i \in I\}$ covering K. This proves that K is compact. □

1.13 Continuous mappings of compact sets

As will soon be clear, the following theorem is a fundamental result in mathematical analysis.

Theorem 1.13.1. *Let $f : X \to Y$ be a continuous mapping between topological spaces X and Y, and let $K \subseteq X$ be a compact subset in X. Then the image $f(K) \subseteq Y$ of K by f is a compact subset in Y.*

Proof. Let $\{V_i \mid i \in I\}$ be an arbitrary open covering of $f(K)$ in Y. We need to prove that we can extract a finite subcovering from it. Since $f : X \to Y$ is continuous, $\{f^{-1}(V_i) \mid i \in I\}$ is an open covering of K in X. Since K is compact, it contains a finite subcovering, say $f^{-1}(V_{i_1}), \ldots, f^{-1}(V_{i_n})$, of K. Clearly, then, $V_{i_1}, \ldots, V_{i_k}$ already covers $f(K)$. Thus we have shown that $\{V_i \mid i \in I\}$ contains a finite subcovering of $f(K)$. Hence $f(K)$ is compact. □

If for $Y = \mathbb{R}$ we combine the above theorem with Theorem 1.11.4 (Heine-Borel), we get the following perfect generalization of the fundamental theorem from classical analysis stating that any continuous function on a closed and bounded interval is bounded; cf. Theorem 1.10.5.

Theorem 1.13.2. *Let $f : X \to \mathbb{R}$ be a continuous function on the topological space X. Then f is bounded in every compact subset K in X. In other words: For every compact subset K in X, there exists a constant $k \in \mathbb{R}^+$ such that $|f(x)| \leq k$ for all $x \in K$.*

Proof. According to Theorem 1.13.1, $f(K)$ is a compact subset in $\mathbb{R}$, and according to Theorem 1.11.4, $f(K)$ is therefore bounded. □

In the proof of Theorem 1.13.2 we did not exploit the full power of Theorem 1.11.4. In fact, the subset $f(K)$ in $\mathbb{R}$ is not only bounded but also closed. Since $f(K)$ is bounded both infimum and supremum of $f(K)$ exist. The numbers $k_1 = \inf f(K)$ and $k_2 = \sup f(K)$ clearly are contact points for $f(K)$ in $\mathbb{R}$, and since $f(K)$ is a closed subset of $\mathbb{R}$, they have to belong to $f(K)$. Therefore there exist points $x_1, x_2 \in K$, such that $f(x_1) = k_1$ and $f(x_2) = k_2$. It is clear, that

$f(x_1)$, respectively $f(x_2)$, is the minimum value, respectively the maximum value that f attains in K. Hence, we have proved the following sharpening of Theorem 1.13.2.

Theorem 1.13.3. *Let $f : X \to \mathbb{R}$ be a continuous function on the topological space X. Then f attains both a maximum value and a minimum value in every compact subset K of X.*

We finish this section by considering the concept of uniform continuity.

Definition 1.13.4. Let (X, d_X) and (Y, d_Y) be metric spaces. A mapping $f : X \to Y$ is said to be *uniformly continuous* in a subset K of X if

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall x, y \in K : \quad d_X(x, y) < \delta \quad \Rightarrow \quad d_Y(f(x), f(y)) < \varepsilon \, .$$

For a given $\varepsilon > 0$, continuity of f in K provides for every $x \in K$ a $\delta_x > 0$ that can be used in a neighbourhood of x only. Or, to be more precise, such that $d_Y(f(x), f(y)) < \varepsilon$ for $y \in K$ with $d_X(x, y) < \delta_x$. Uniform continuity on the other hand provides a $\delta > 0$ that can be used everywhere in K.

Concerning the relationship between continuity and uniform continuity we have the following useful result.

Theorem 1.13.5. *Let $f : X \to Y$ be a continuous mapping between metric spaces (X, d_X) and (Y, d_Y). Then f is uniformly continuous in every compact subset K of X.*

Proof. Let K be a compact subset of X. For any given $\varepsilon > 0$ we shall provide a $\delta > 0$, which can be used all over K as required by definition 1.13.4.

So let $\varepsilon > 0$ be given. Since f is continuous, we can for every $x \in K$ choose a $\delta_x > 0$, such that $d_Y(f(x), f(y)) < \varepsilon/2$ for all $y \in K$ with $d_X(x, y) < \delta_x$. The system of open balls $\{B_{\delta_x/2}(x) \mid x \in K\}$ is an open covering of K. Since K is compact, this covering contains a finite subcovering, and hence there exist finitely many points $x_1, \ldots, x_n$ in K, such that $B_{\delta_{x_1}/2}(x_1), \ldots, B_{\delta_{x_n}/2}(x_n)$ cover K. Now, set

$$\delta = \min\{\delta_{x_1}/2, \ldots, \delta_{x_n}/2\} \, .$$

We claim that this $\delta > 0$ works. Indeed, consider any pair of points $x, y \in K$ with $d_X(x, y) < \delta$. For at least one of the finitely many points $x_1, \ldots, x_n$, say x_i, we have $d_X(x_i, x) < \delta_{x_i}/2$. The triangle inequality then gives

$$d_X(x_i, y) \leq d_X(x_i, x) + d_X(x, y) < \delta_{x_i} .$$

From the way δ_{x_i} was determined, we have

$$d_Y(f(x_i), f(x)) < \varepsilon/2 \quad \text{and} \quad d_Y(f(x_i), f(y)) < \varepsilon/2.$$

The triangle inequality now immediately implies

$$d_Y(f(x), f(y)) < \varepsilon,$$

which is all we had to show. This completes the proof. □

1.14 Homeomorphisms

Recall that a mapping $f : X \to Y$ between sets X and Y is said to be *injective* if for any pair of points $x_1, x_2 \in X$,

$$f(x_1) = f(x_2) \quad \Rightarrow \quad x_1 = x_2,$$

or, equivalently,

$$x_1 \neq x_2 \quad \Rightarrow \quad f(x_1) \neq f(x_2).$$

The mapping f is said to be *surjective* if for every point $y \in Y$ there exists at least one point $x \in X$ such that $f(x) = y$.

If f is both injective and surjective, f is said to be *bijective*. A bijective mapping $f : X \to Y$ has an *inverse mapping*, which we denote by $f^{-1} : Y \to X$; for $y \in Y$, define $f^{-1}(y) = x$ as the unique $x \in X$ for which $f(x) = y$.

Let $1_X : X \to X$, respectively $1_Y : Y \to Y$, denote the identity mapping (the identity) of X, respectively Y, defined by $1_X(x) = x$ for all $x \in X$, respectively $1_Y(y) = y$ for all $y \in Y$. Bijectivity can then be given the following alternative formulation.

Lemma 1.14.1. *A mapping $f : X \to Y$ between sets X and Y is bijective if and only if there exists a mapping $g : Y \to X$ such that*

$$g \circ f = 1_X \quad \text{and} \quad f \circ g = 1_Y.$$

If this is the case, then $g = f^{-1}$.

Proof. The condition $g \circ f = 1_X$ ensures that f is injective, and the condition $f \circ g = 1_Y$ ensures that f is surjective. □

Now, let us return to topological spaces and continuous mappings.

Definition 1.14.2. Let X and Y be topological spaces. A *homeomorphism* is a bijective mapping $f : X \to Y$ for which both f and the inverse mapping f^{-1} are continuous.

Two topological spaces X and Y are said to be *homeomorphic*, or, *topologically equivalent*, if there exists a homeomorphism between them.

We note that a continuous mapping $f : X \to Y$ is a homeomorphism if and only if there exists a continuous mapping $g : Y \to X$ such that

$$g \circ f = 1_X \quad \text{and} \quad f \circ g = 1_Y.$$

We also note that if $f : X \to Y$ is a homeomorphism then it defines a bijective correspondence between the open sets in X and the open sets in Y. In particular, we would like to stress that the continuity of f^{-1} implies that if U is an open set in X then its image $f(U)$ is an open set in Y. Therefore the topological structures in two homeomorphic topological spaces X and Y are indistinguishable. In other words, homeomorphic topological spaces have the same topological properties. For instance, either both spaces satisfy the Hausdorff Axiom or neither of them do.

Often, it is difficult to determine the inverse mapping $f^{-1} : Y \to X$ of a bijective mapping $f : X \to Y$ explicitly, and hence it may be difficult to check whether it is continuous. In the theorem below, however, we show that if X is compact and Y is a Hausdorff space, then the inverse mapping $f^{-1} : Y \to X$ for a continuous, bijective mapping $f : X \to Y$ is automatically continuous. This result is very useful.

Yet, there are important situations, for example in connection with function spaces, in which a mapping is continuous and bijective but the inverse mapping fails to be continuous. One gets a trivial example of this phenomenon by considering $X = \mathbb{R}$ with the discrete topology, and $Y = \mathbb{R}$ with the usual topology. Then the identity map $1_{\mathbb{R}}$ on $\mathbb{R}$ is continuous, when considered as a mapping from X to Y, but it is not continuous when considered as a mapping from Y to X.

Theorem 1.14.3. *Let X and Y be topological spaces, where X is compact and Y is Hausdorff. If $f : X \to Y$ is a continuous, bijective mapping then the inverse mapping $f^{-1} : Y \to X$ is continuous. In this case, $f : X \to Y$ is therefore a homeomorphism.*

Proof. We shall prove that f^{-1} is continuous. According to Theorem 1.7.3, it suffices to show that $(f^{-1})^{-1}(A) = f(A)$ is a closed set in Y for any closed set A in X. Therefore, let A be a closed set in X. Since X is compact, it follows by Theorem 1.12.5, that A is a compact subset of X. Then Theorem 1.13.1 implies that $f(A)$ is a compact subset of Y since f is continuous. Finally, $f(A)$ is therefore closed in the Hausdorff space Y by Theorem 1.11.1. This completes the proof. □

The following example demonstrates a typical way of using Theorem 1.14.3.

Example 1.14.4. Let $S^1 = \{(\cos\theta, \sin\theta) \mid \theta \in \mathbb{R}\}$ be the unit circle in the plane equipped with the induced topology. We shall give an alternative description of S^1.

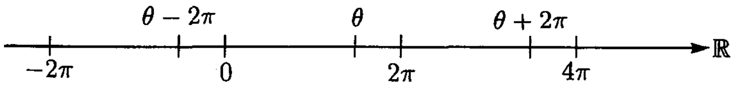
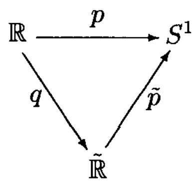


Figure 1.9

In $\mathbb{R}$ we introduce the relation $\sim$ by setting $\theta \sim \theta'$, if $\theta - \theta'$ is an integral multiple of 2π; cf. Figure 1.9.

It is easy to show that $\sim$ is an equivalence relation in $\mathbb{R}$. This way the points of $\mathbb{R}$ are divided into equivalence classes of points, which are separated from each other by multiples of 2π. Note that every equivalence class has exactly one representative in the interval $[0, 2\pi[$. Let $\tilde{\mathbb{R}}$ denote the set of equivalence classes in $\mathbb{R}$ under $\sim$, and let $q : \mathbb{R} \to \tilde{\mathbb{R}}$ be the mapping which maps $\theta \in \mathbb{R}$ into it equivalence class $q(\theta) \in \tilde{\mathbb{R}}$. We equip $\tilde{\mathbb{R}}$ with the quotient topology induced from $\mathbb{R}$ by the mapping q. Our objective is to show that $\tilde{\mathbb{R}}$ and S^1 are homeomorphic topological spaces.

Define the mapping $p : \mathbb{R} \to S^1$ by $p(\theta) = (\cos\theta, \sin\theta)$ for $\theta \in \mathbb{R}$. Since Cosine and Sine are periodic functions of period 2π, it is easy to see that $p : \mathbb{R} \to S^1$ induces a mapping $\tilde{p} : \tilde{\mathbb{R}} \to S^1$ such that $p = \tilde{p} \circ q$. The situation is illustrated by the diagram:



Clearly $\tilde{p}$ is bijective. Since $\tilde{\mathbb{R}}$ is equipped with the quotient topology induced from $\mathbb{R}$ by the mapping q, and since p is continuous, it is easy to show that $\tilde{p}$ is also continuous. Since $\tilde{\mathbb{R}}$ is compact (because $\tilde{\mathbb{R}}$ is the image of the closed and bounded interval $[0, 2\pi]$ under the continuous mapping q), and since S^1 is a Hausdorff space (it is even metric), it follows by Theorem 1.14.3 that $\tilde{p}$ is a homeomorphism.

In other words, the topological spaces S^1 and $\tilde{\mathbb{R}}$ are homeomorphic.

1.15 Connected sets

Informally speaking, a topological space is said to be connected if it does not consist of several disjoint pieces. However, we need to be more precise to make this into a rigorous definition. First a small preparation.

Lemma 1.15.1. *Let S be a topological space. Then the following statements are equivalent:*

(1) *$\emptyset$ and S are the only subsets of S which are both open and closed.*

(2) *If U_1 and U_2 are open sets in S which partition S into disjoint pieces, i.e. $S = U_1 \cup U_2$ and $U_1 \cap U_2 = \emptyset$, then either $U_1 = \emptyset$ or $U_2 = \emptyset$.*

(3) *If A_1 and A_2 are closed sets in S which partition S into disjoint pieces, i.e. $S = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$, then either $A_1 = \emptyset$ or $A_2 = \emptyset$.*

Proof. First we prove (1)$\Rightarrow$(2). Assume that U_1 and U_2 are open sets in S which partition S into disjoint pieces. Since $U_1 = S \setminus U_2$, we see that U_1 is also closed. From (1) it follows that either $U_1 = \emptyset$ or $U_1 = S$, but in the latter case $U_2 = S \setminus U_1 = \emptyset$.

Next we prove (2)$\Rightarrow$(3). Assume that A_1 and A_2 are closed sets in S which partition S into disjoint pieces. If we set $U_1 = S \setminus A_1$ and $U_2 = S \setminus A_2$ it is clear that U_1 and U_2 are open sets in S which partition S into disjoint pieces. According to (2) either $U_1 = \emptyset$, in which case $A_2 = \emptyset$, or $U_2 = \emptyset$, in which case $A_1 = \emptyset$.

Finally we prove (3)$\Rightarrow$(1). Assume that W is a subset of S which is both open and closed. If we set $A_1 = W$ and $A_2 = S \setminus W$, then it is clear that A_1 and A_2 are closed sets in S which partition S into disjoint pieces. According to (3), we have that either $A_1 = \emptyset$, thus $W = \emptyset$, or $A_2 = \emptyset$, thus $W = S$.

Since we have now proved all implications in the cycle (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1), the proof is completed. □

Definition 1.15.2. A topological space S, which satisfies one of the equivalent conditions (1), (2) or (3) in Lemma 1.15.1, and therefore all of them, is called *connected.*

A subset T in S is called connected if T is connected, when equipped with the topology induced from S.

After this slightly abstract definition the following result must be reassuring:

Lemma 1.15.3. *A closed and bounded interval $[a, b]$ in $\mathbb{R}$ is connected.*

Proof. First, note that since $[a, b]$ is a closed set in $\mathbb{R}$, any closed subset in $[a, b]$ equipped with the subspace topology, is also a closed set in $\mathbb{R}$.

The proof is by contradiction. Assume that $[a,b]$ is not connected. Then we can find closed sets A_1 and A_2 in $\mathbb{R}$, such that $[a,b] = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, $A_1 \neq \emptyset$, and $A_2 \neq \emptyset$. Possibly after interchanging A_1 and A_2 we may assume that $b \in A_2$. Since A_1 is bounded from above, the number $c = \sup A_1$ exists. Now A_1 is closed, and hence $c \in A_1$. Since b is an upper bound for A_1, and $b \notin A_1$, we have $c < b$. But then $]c,b] \subseteq A_2$, since $c = \sup A_1$. Hence any open interval around c contains points from A_2. This shows that c is a contact point of A_2, and therefore $c \in A_2$, since A_2 is closed. Altogether $c \in A_1 \cap A_2$, which clearly contradicts that $A_1 \cap A_2 = \emptyset$. The assumption that $[a,b]$ is not connected must therefore be rejected, and the lemma follows. □

Connectivity is - like compactness - a property of topological spaces that is preserved by continuous mappings. This is the content of the following theorem.

Theorem 1.15.4. *Let $f : X \to Y$ be a continuous mapping between topological spaces X and Y, and let $T \subseteq X$ be a connected subset in X. Then the image $f(T) \subseteq Y$ of T under f is a connected subset in Y.*

Proof. A subset in a topological space is connected, if it is connected when considered as a topological space equipped with the subspace topology. Hence it suffices to consider the topological spaces $X' = T$ and $Y' = f(T)$. The restriction of f to X' defines a continuous, surjective mapping $f' : X' \to Y'$ of X' onto Y'. Therefore it suffices to prove that if X' is connected then Y' is also connected.

Now the proof is by contradiction. Assume therefore that Y' is not connected. Then there exist open sets V_1' and V_2' in Y', such that $Y' = V_1' \cup V_2'$, $V_1' \cap V_2' = \emptyset$, $V_1' \neq \emptyset$, and $V_2' \neq \emptyset$. By the continuity and surjectivity of $f' : X' \to Y'$, the sets $U_1' = f'^{-1}(V_1')$ and $U_2' = f'^{-1}(V_2')$ constitutes a similar partitioning of X' into two disjoint, non-empty, open sets. This contradicts X' being connected. But then Y' must be connected, as was to be proved. □

The following is perhaps a more intuitive concept of connectivity.

Definition 1.15.5. A topological space S is called *pathwise connected* if every pair of points $x, y \in S$ can be joined by a curve (path); that is, if there exists a continuous mapping $\varphi : [0,1] \to S$, such that $\varphi(0) = x$ and $\varphi(1) = y$.

Theorem 1.15.6. *A pathwise connected topological space S is connected.*

Proof. Assume that S is pathwise connected. Let $W \subseteq S$ be a subset of S which is both open and closed, and assume that $W \neq \emptyset$. According to the first of the equivalent definitions of connectivity, we have to prove that $W = S$.

Choose a fixed point $x \in W$, and let $y \in S$ be an arbitrary point in S. Since S is pathwise connected, there exists a continuous mapping $\varphi : [0,1] \to S$, such that $\varphi(0) = x$ and $\varphi(1) = y$. By the continuity of φ, the preimage $\varphi^{-1}(W)$

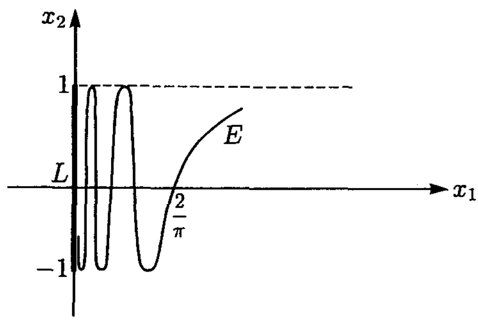


Figure 1.10

of W under φ is both open and closed in $[0,1]$. Since $0 \in \varphi^{-1}(W)$, the set $\varphi^{-1}(W) \neq \emptyset$, and therefore $\varphi^{-1}(W) = [0,1]$ since $[0,1]$ is connected by Lemma 1.15.3. Hence $y = \varphi(1) \in W$. This shows that $W = S$, and therefore that S is connected. □

In general, a connected topological space need not be pathwise connected.

Example 1.15.7. Consider in $\mathbb{R}^2$ the subset $T = L \cup E$, where

$$L = \left\{(x_1, x_2) \in \mathbb{R}^2 \;\middle|\; x_1 = 0,\; -1 \leq x_2 \leq 1\right\},$$

and

$$E = \left\{(x_1, \cos(1/x_1)) \in \mathbb{R}^2 \;\middle|\; x_1 > 0\right\}.$$

Intuitively it is clear that $T = L \cup E$ is not pathwise connected, but T is actually connected.

That T is connected can be seen by noticing that T is exactly the closure of the pathwise connected, and therefore also connected, set E. Then the rest follows from a general result which states: *In an arbitrary topological space S, the closure $\overline{E}$ of a connected subset E in S is also connected.* We leave the proof of this statement to the reader.

As an extension of the result stated in Lemma 1.15.3, we can now characterize the connected subsets of the real axis.

Theorem 1.15.8. *A subset T in $\mathbb{R}$ is connected if and only if it is an interval in the extended sense (i.e. half-lines and $\mathbb{R}$ itself are included).*

Proof. First, assume that T is connected. We shall prove that T is an interval in the extended sense. Assume that this is not the case. Then we can find points $x, y, z \in \mathbb{R}$, such that $x < y < z$, where $x, z \in T$ but $y \notin T$. Consider the sets

$$U_1 =]-\infty, y[\cap T \quad \text{and} \quad U_2 =]y, +\infty[\cap T.$$

Since T is equipped with the topology induced from $\mathbb{R}$, the sets U_1 and U_2 are open sets in T with $T = U_1 \cup U_2$ and $U_1 \cap U_2 = \emptyset$. From $x \in U_1$ and $z \in U_2$ it follows that we have a partitioning of T into disjoint, non-empty, open sets, which contradicts the connectivity of T. Therefore T is an interval in the extended sense.

Next, assume that T is an interval in the extended sense. Then T is clearly pathwise connected, and hence connected. □

As a result of our efforts, we can now prove the following perfect generalization of the well-known and useful theorem from classical analysis stating that under a continuous function $f : [a, b] \to \mathbb{R}$, an interval $[a, b]$ is mapped onto an interval $[c, d]$.

Theorem 1.15.9. *Let $f : S \to \mathbb{R}$ be a continuous function defined on a compact, connected topological space S. Then the image $f(S)$ of S under f is a closed and bounded interval $[c, d]$ in $\mathbb{R}$.*

Proof. Using Theorem 1.15.4 it follows by the connectivity of S that $f(S)$ is a connected subset in $\mathbb{R}$, and hence according to Theorem 1.15.8 an interval in the extended sense.

Since S is compact, $f(S)$ is also compact according to Theorem 1.13.1, and thus by Theorem 1.11.4, $f(S)$ is a closed and bounded subset of $\mathbb{R}$.

Altogether, $f(S)$ is a closed and bounded interval $[c, d]$ in $\mathbb{R}$ as asserted. □