1 The real numbers as a metric space

In this section we recall some basic properties of the real numbers and real-valued functions. The familiar notions of limit, convergence, and continuity are expressed in terms of the metric or distance function. This allows us to define the above mentioned notions in the more general context of metric spaces, which we will need later on.

Good references for this and several of the following sections are
G.F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.
J.E. Marsden, *Elementary Classical Analysis*, 2nd ed., W.H. Freeman and Company, New York, 1993.

The majority of the material of Sections 1 and 2 is also covered in
W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York, 1964

1.1 Definition of Metric

On a first reading, substitute $\mathbb{R}$ for the arbitrary set X in the following definition and subsequent discussion. This week we will discuss almost exclusively $\mathbb{R}$ and $\mathbb{R}^n$ (Euclidean space), but more general (and usually infinite-dimensional) metric spaces will follow soon.

The following is the definition of *metric*, also called *distance function*, on an arbitrary set X.

Definition 1.1 (metric) *A metric on X is a function $d : X \times X \to \mathbb{R}$, with the following properties:*
(M1) $d(x, y) \geq 0$ for all $x, y \in X$, and $d(x, x) = 0$.
(M2) If $d(x, y) = 0$, then $x = y$.
(M3) $d(x, y) = d(y, x)$, for all $x, y \in X$.
(M4) $d(x, y) \leq d(x, z) + d(z, y)$, for all $x, y, z \in X$.

In English the four defining properties read:
M1: Distances are non-negative, and the distance between a point and itself vanishes
M2: The only point at distance 0 from x is x itself.
M3: The distance is a *symmetric* function.
M4: This is the so-called *Triangle Inequality* . When applied to the Euclidean plane, i.e., $X = \mathbb{R}^2$ with the usual distance function, this inequality expresses the fact that the length of any side of a triangle is less than the sum of the lengths of the two other sides. M2 is also called *strict positivity* or *non-degeneracy.*

X, equipped with a metric d, is called a metric space, often denoted by (X, d), especially when it is important to specify which metric is being considered.

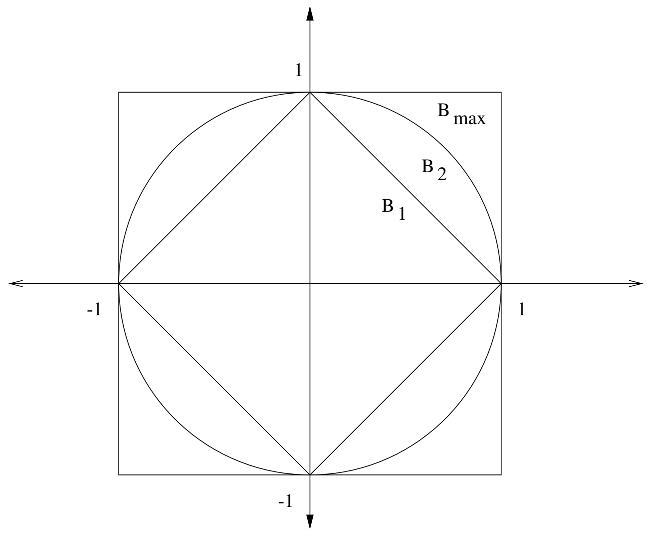

Figure 1.1: The unit ball in $\mathbb{R}^2$ for the Euclidean norm (B_2), the sum norm (B_1), and the maximum norm ($B_{\max}$).

Examples:

1. $X = \mathbb{R}$, and $d(x, y) = |x - y|$.

2. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and the Euclidean metric given by
$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$
Note: the distance of $x \in \mathbb{R}^n$ to $0 \in \mathbb{R}^n$ (the origin), is called the length or *norm* of x, denoted by $\|x\|$. The obvious relation between the norm and the metric is then
$$d(x, y) = \|x - y\|$$

3. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and
$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n|$$
Note: the corresponding norm $\|x\| = d(0, x)$ is called the *sum norm* or the L^1 norm .

4. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and
$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\}$$
Note: the corresponding norm $\|x\| = d(0, x)$ is usually called the *maximum norm* or the L^∞ norm. The unit balls in $\mathbb{R}^2$ for each of the metrics defined in Examples 2-4 are shown in Figure 1.1.

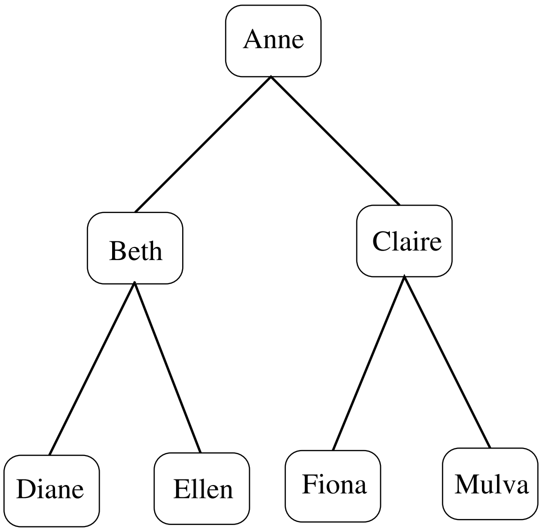

Figure 1.2: An example of a family tree as used in the definition of the ultrametric defined in (1.1).

5. Let X be a set people of one and the same generation and with a common ancestor, e.g. all greatgrandchildren of greatgrandmother $\mathcal{M}$. A metric d can then be defined on this set by putting $d(x, y) =$ the number of generations one has to go back to find the first common ancestor, along female lines, of any two individuals x and y. E.g., the distance between two sisters is 1. Then

$$d(x, y) \leq \max\{d(x, z), d(z, y)\} \quad \text{for all } x, y, z \in X \tag{1.1}$$

   and a fortiori M4 holds. See Figure 1.2 for an illustration. (1.1) is stronger than the triangle inequality. When it holds the metric d is called an *ultrametric*. In taxonomy ultrametrics have been used to characterize genetic proximity of species.

6. Let X be the set of all N-letter words in an alphabet A with k characters, i.e.,

$$X = \{(x_1, x_2, \ldots, x_N) \mid x_i \in A\}$$

   and $A = \{a_1, a_2, \ldots, a_k\}$. A natural metric on X is then defined by

$$d(x, y) = \#\{i \mid x_i \neq y_i\}$$

7. The metric spaces most important for this course are infinite-dimensional ones, of which there is a large variety. Before we can start introducing those we need to review just a little more real analysis.

Figure 1.3: A sequence x_n converging to x.

1.2 When distances tend to zero: convergence

1.2.1 Sequences

For concreteness let us consider again the real numbers with their usual metric. All definitions, however, do carry over to arbitrary metric spaces. Just replace $\mathbb{R}$ by X, $|x-y|$ by $d(x,y)$, etc.

Let (x_n) be a *sequence* of real numbers, i.e., for each $n \in \mathbb{N}, x_n \in \mathbb{R}$ (other common notations are $\{x_n\}, (x_n)_{n\geq 1}, \{x_n\}_{n=0}^{\infty}$ etc.

Definition 1.2 (convergence) *The sequence of real numbers (x_n) converges to $x \in \mathbb{R}$ iff for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n \geq N$, $|x_n - x| < \epsilon$.*

Note that in this definition N depends on ϵ, and you could write $N(\epsilon)$ to make this dependence explicit. The point x is called the *limit* of (x_n). Common notations expressing that (x_n) converges to x are:

$$\lim_{n\to\infty} x_n = x, \quad x_n \to x$$

Sequences that become eventually larger than any number do not converge according to the definition given above. They are usually called *divergent*. But it is sometimes more useful to consider them as converging to $+\infty$, i.e., $x_n \to +\infty$ iff for all $M \in \mathbb{R}$, there is $N \in \mathbb{N}$, such that for all $n \geq N$, $x_n > M$. And similarly one defines $x_n \to -\infty$ iff for all $M \in \mathbb{R}$, there is $N \in \mathbb{N}$, such that for all $n \geq N$, $x_n < M$.

For an arbitrary metric space (X,d), convergence of a sequence (x_n), $x_n \in X$, to $x \in X$ means that for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n \geq N$, $d(x_n, x) < \epsilon$. Figure 1.3 shows a convergent sequence in $\mathbb{R}^2$, considered with the Euclidean metric (or any one of the other metrics defined in Examples 3 and 4).

Next, we define the concept of a *Cauchy* sequence.

Definition 1.3 (Cauchy sequence) *(x_n) is a Cauchy sequence iff for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n, m \geq N$, $|x_n - x_m| < \epsilon$.*

Cauchy sequences are sequences that *ought* to converge. The following propositon says that, in general, a necessary condition for any sequence to converge is that it is Cauchy.

Proposition 1.4 *If (x_n) converges to $x \in \mathbb{R}$, then (x_n) is Cauchy.*

This proposition holds for sequences in any metric space (X, d). A natural question now is whether the converse is also true: does every Cauchy sequence have a limit it converges to? For real sequences the answer is *yes*, but this is not so in general. The property that every Cauchy sequence in a metric space (X, d) converges to a limit *in* X, singles out a particularly interesting class of metric spaces. Such (X, d) are called *complete* . So, $\mathbb{R}$ with the usual metric is complete, but the rational numbers, $\mathbb{Q}$, considered as a subset of $\mathbb{R}$ and with the same distance function, is *not* complete. In fact, one way of constructing the set of real numbers is by considering all Cauchy sequences in $\mathbb{Q}$ and grouping those together in equivalence classes of sequences that ought to have the same limit. We will discuss the notion of completeness and its applications in Chapter 4.

Exercise 1.1 *Prove that $\mathbb{Q}$ with the usual distance function is not complete as a metric space. (Hint: find a sequence of rational numbers that is Cauchy, and hence converges in $\mathbb{R}$, but with an irrational limit.)*

Exercise 1.2 *Starting from the fact that $\mathbb{R}$ equipped with its usual distance function is complete, prove that*
a) *$\mathbb{R}^n$ with the Euclidean norm is a complete metric space,*
b) *$\mathbb{R}^n$ with the sum norm or the maximum norm is also complete.*

Another natural question is whether a sequence could converge to two different limits. In a metric space, the answer is *no*, in particular due to property M2 of the metric.

Proposition 1.5 *If $x_n \to x$ and $x_n \to y$, then $x = y$*

A *subsequence* of a sequence (x_n) is a sequence of the form $(x_{n_k})_{k=1}^{\infty}$, where for each $k \in \mathbb{N}$, $n_k \in \mathbb{N}$ and $n_k < n_{k+1}$, for all k. I.e., $k \mapsto n_k$ is a strictly increasing function from the set of natural numbers to itself. Clearly, if (x_n) is a convergent sequence with limit x, then all its subsequences also converge to x.

1.2.2 Upper and lower bound, inf and sup, lim inf and lim sup

There is a natural order relation on $\mathbb{R}$ that is nicely related to the metric structure. First, we define the notion of *upper* and *lower bound* for a subset A of $\mathbb{R}$.

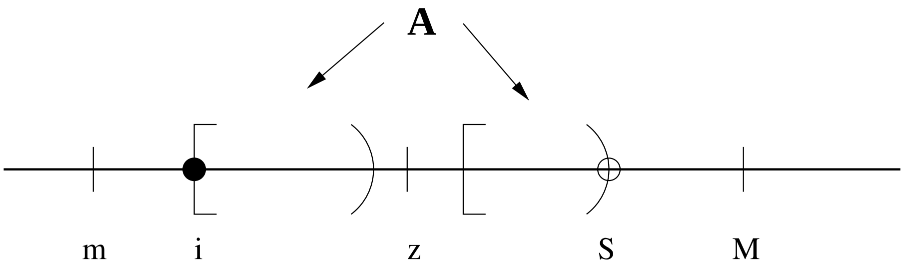

Figure 1.4: M is an upper bound for A and m is lower bound for A. z is neither an upper bound nor a lower bound. S is the supremum of A (but does not belong to A). i is the infimum (and, as $i \in A$, also the minimum) of A.

Definition 1.6 (upper and lower bound) *Let $A \subset \mathbb{R}$. $M \in \mathbb{R}$ is an upper bound for A iff $x \leq M$, for all $x \in A$, and $m \in \mathbb{R}$ is a lower bound for A iff $m \leq x$, for all $x \in A$. The set A is called bounded iff both an upper and a lower bound for it exist.*

If A has an upper bound M, then A has many upper bounds. E.g, any $M' \geq M$ is also an upper bound. The *least upper bound* for A is called the *supremum* of A, denoted by $\sup A$. I.e., $\sup A$ is an upper bound with the property that if M is any upper bound for A, then $M \geq \sup A$.

Similarly one defines $\inf A$, the *infimum* of A, as the *greatest lower bound* for A. I.e., $\inf A$ is a lower bound with the property that if m is any lower bound for A, then $m \leq \inf A$.

By convention we set $\sup A = +\infty$ if A does not have an upper bound, i.e., if A contains arbitrarily large real numbers. In the same way, one sets $\inf A = -\infty$ if A does not have a lower bound.

Obviously, for each $A \neq \emptyset$, there can be only one least upper bound (greatest upper bound). This is due to the fact that $\mathbb{R}$ is *totally* ordered, i.e., for all $x, y \in \mathbb{R}$, $x \leq y$ or $y \leq x$. One also has $\inf A \leq \sup A$, for all $A \neq \emptyset$. For $A = \emptyset$ the convention $\inf A = +\infty$ and $\sup A = -\infty$ is sometimes useful.

That $\sup A$ ($\inf A$) exists for any set with an upper (lower) bound, is a consequence of (and actually equivalent to) the completeness of $\mathbb{R}$.

If $\sup A \in A$, $\sup A$ is also the *maximum* of A, defined as the largest element of A, and often denoted by $\max A$. If $\inf A \in A$, $\inf A$ is also the *minimum* of A, defined as the smallest element of A, and often denoted by $\min A$. Note that every bounded set has a supremum and an infimum, but that it does not necessarily have a maximum or a minimum.

The concepts of maxiumum, minimum, supremum, infimum, and upper and lower bound, are illustrated in Figure 1.4.

The following two quantities can again be defined in an arbritrary metric space (X, d).

For any $A \subset X$, the *diameter* of A is defined by

$$\text{diam}A = \sup\{d(x,y) \mid x, y \in A\}$$

For a subset $A \subset X$ and a point $x \in X$ the distance from x to A is naturally defined by

$$\text{dist}(x, A) \equiv d(x, A) = \inf\{d(x,y) \mid y \in A\}$$

Note that the distance function dist is *not* a metric on the power set of X.

Examples:

1. $A = \{1, 1/2, 1/3, 1/4, \ldots\}$. Then $\sup A = 1 = \max A$ and $\inf A = 0$. Note that, in this example, A does not have a minimum.

2. With A as given in 1), $\text{diam}A = 1$. Note that $d(0, A) = 0$, although $0 \notin A$.

In order to define lim inf and lim sup we first observe that monotone sequences always converge (allowing for convergence to $\pm\infty$). Indeed, if (x_n) is *monotone increasing*, i.e., $x_n \le x_{n+1}$, for all n, and (y_n) is *monotone decreasing*, i.e., $x_n \ge x_{n+1}$, for all n, then

$$\lim_{n\to\infty} x_n = \sup\{x_n\} \tag{1.2}$$
$$\lim_{n\to\infty} y_n = \inf\{y_n\} \tag{1.3}$$

lim inf (*limes inferior*) and lim sup (*limes superior*) can now be defined for an arbitrary sequence (x_n), as the following two limits:

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k \mid k \ge n\} \tag{1.4}$$
$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k \mid k \ge n\} \tag{1.5}$$

Another common notation is: $\liminf = \underline{\lim}, \limsup = \overline{\lim}$.

Exercise 1.3 *Let (x_n) be a sequence of real numbers. A point $c \in \mathbb{R} \cup \{\pm\infty\}$ is called a* cluster point *of (x_n) if there is a convergent subsequence of (x_n) with limit c. Let C denote the set of all cluster points of (x_n). Prove that*

$$\limsup x_n = \max C, \quad \textit{and} \quad \liminf x_n = \min C \quad .$$

Exercise 1.4 *Let (x_n) be a sequence of real numbers. Prove that, for every $\epsilon > 0$, and every N, there exists $n_1, n_2 > N$, such that*

$$x_{n_1} + \epsilon \ge \limsup_m x_m, \quad x_{n_2} - \epsilon \le \liminf_m x_m$$

In contrast to the limit, the lim inf and lim sup always exist. Clearly, one always has $\liminf_n x_n \leq \limsup_n x_n$. Note that a sequence (x_n) converges to a limit x iff $\liminf_n x_n = \limsup_n x_n = x$.

Let A be an arbitrary set, and $x_\alpha \in \mathbb{R}$, for all $\alpha \in A$. Then $\sup\{x_\alpha \mid \alpha \in A\}$ is usually abbreviated to $\sup_{\alpha \in A} x_\alpha$, or, when this does not lead to confusion, by $\sup_\alpha x_\alpha$. Now consider a family of real numbers indexed by $n \in \mathbb{N}$ and $\alpha \in A$: $x_{n,\alpha}$. The following two relations are very useful and easy to prove:

$$\limsup_{n\to\infty} \inf_\alpha x_{n,\alpha} \leq \inf_\alpha \limsup_{n\to\infty} x_{n,\alpha} \tag{1.6}$$
$$\sup_\alpha \liminf_{n\to\infty} x_{n,\alpha} \leq \liminf_{n\to\infty} \sup_\alpha x_{n,\alpha} \tag{1.7}$$

Exercise 1.5 *Prove the inequalities (1.6) and (1.7)*

Finally, we mention the notion of *subsequence*. A sequence (y_k) is called a subsequence of the sequence (x_n), if there is a strictly increasing sequence of positive integers n_k such that $y_k = x_{n_k}$, for all k.

1.3 Series

Let (x_n) be a sequence in $\mathbb{R}$. The sequence of *partial sums* of (x_n) is the new sequence (s_n) defined by

$$s_n = \sum_{k \leq n} x_k$$

If (s_n) converges to a limit $s \in \mathbb{R}$, we say that the *series* $\sum_{n=1}^\infty x_n$ converges to s. If (s_n) does not converge the series is called *divergent*.

A series $\sum_{n=1}^\infty x_n$ is said to be *absolutely convergent* iff the series of absolute values $\sum_{n=1}^\infty |x_n|$ converges. Absolute convergence implies convergence, but not vice versa.

Exercise 1.6 *Show that the series*

$$\sum_{n=1}^\infty (-1)^n \frac{1}{n}$$

is not absolutely convergent. Show that by permuting the terms of this series one can obtain series with different limits.

A useful feature of absolutely convergent series is that any series obtained from it by a permutation of its terms, is also convergent and has the same limit. For series of real numbers the property that any permutation converges to the same limit (this property is called *unconditional convergence*) is in fact equivalent to absolute convergence. This is no longer true for series in infinite-dimensional spaces.

1.4 Open and closed sets

Consider a metric space (X, d), e.g., $\mathbb{R}^n$. The *open ball*, $B_r(x_0)$, with radius $r > 0$ and center $x_0 \in X$ is defined by

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

The *closed ball*, $\overline{B}_r(x_0)$, is defined by

$$\overline{B}_r(x_0) = \{x \in X \mid d(x_0, x) \leq r\}$$

The basic concept of (set) topology is the notion of *open set*.

Definition 1.7 (open and closed sets) *A subset $G \subset X$ is open iff for all $x \in G$ there exists $r > 0$ such that $B_r(x) \subset G$.*
A subset $F \subset X$ is closed iff $F^c \equiv X \setminus F$ is open.

Proposition 1.8 *A subset F of a metric space is closed iff all convergent sequences of elements in F converge to a limit in F, i.e., if $x_n \to x$, and $x_n \in F$, for all n, then $x \in F$.*

A set $A \subset X$ is called bounded if it is contained in a ball $B_r(x_0)$, with $r < +\infty$. The following properties are easy to prove:

1. The empty set $\emptyset$ and X itself are open *and* closed.
2. The intersection $G_1 \cap G_2$ of two open sets $G_1, G_2 \subset X$ is open. Consequently, the intersection of any finite family of open sets is also open.
3. The union of an arbitrary family (finite or infinite) of open sets is open.
4. The union $G_1 \cup G_2$ of two closed sets $G_1, G_2 \subset X$ is closed, and hence any finite family of closed sets has a closed union.
5. The intersection of an arbitrary family (finite or infinite) of closed sets is closed.

Exercise 1.7 *Prove the properties 1-5.*

The closure $\overline{A}$ of a set $A \subset \mathbb{R}^n$, is, by definition, the smallest closed set containing A. The last property above implies that such a set indeed exists. In any metric space the closure of a subset A can be obtained by adding to A all limits of converging sequences of elements of A. I.e,

$$\overline{A} = \{x \in X \mid \text{there exist } x_n \in A, x_n \to x\}$$

Therefore, $\overline{A}$ is sometimes *defined* as the set of *points of adherence* of A. By definition, $x \in X$ is a *point of adherence* of A if every open set containing x also contains points of A.

For any $x \in X$ a *neighborhood* of x is a subset $V \subset X$ such that there is an open set $G \subset V$ containing x. Equivalently, V is a neighborhood of x if there is a ball $B_r(x) \subset V$, for some $r > 0$.

A set $A \subset X$ is called *dense* if $\overline{A} = X$.

1.5 Countably and uncountably infinite sets

This is a brief aside on infinite sets. Our only aim here is to clarify the distinction between countably infinite and uncountably infinite sets.

A set A is finite, i.e., has a finite number of elements, iff there is a bijection (a map that is one-to-one and onto), from A to one of the sets $\emptyset$ or $\{1, 2, \ldots, n\}$. Of course, n is the number of elements of A. If a set is not finite, it is called *infinite*.

A set A is *countable* if it is either finite or if there is a bijection from A to $\mathbb{N}$. In the latter case it is called *countably infinite*, i.e., in order to enumerate all the elements of A, you need an infinite number of natural numbers.

Sets that are not countable are *uncountable*. To label all the elements of an uncountable set, the natural numbers do not suffice, i.e., they cannot be enumerated.

Examples:

1. Any infinite subset of $\mathbb{N}$ (or any other countable set) is countably infinite. So, e.g., the even natural numbers form a countably infinite set.

2. $\mathbb{Z}$ is countably infinite.

3. $\mathbb{Q}$ is countably infinite. Do you know an enumeration of the rational numbers?

4. $\mathbb{R}$ is uncountably infinite. Do you know how to prove this?

1.6 Compactness

The full importance of the *compact sets* will become clear only in an infinite-dimensional setting, and we will have ample opportunity to see how useful they are later. It is nevertheless interesting to start with a finite-dimensional warmup.

Definition 1.9 *A subset K of a metric space X is called compact iff every sequence (x_n) of elements of K has a convergent subsequence with a limit $x \in K$.*

Sets satisfying the definition above are sometimes called *sequentially compact* in order to make a distinction with the more general definition of compactness in the framework of general topological spaces (of which the metric spaces are a special case). Using the general definition here would force us to also prove that for metric spaces compactness is equivalent to sequential compactness, which is the more useful notion for the applications we have in mind. Therefore, it is more economic not to go beyond the scope of metric spaces now. We will discuss the more general situation later.

Exercise 1.8 *Prove that every compact set is closed.*

The compact subsets of $\mathbb{R}^n$ have a simple and practical characterization given by the following theorem.

Theorem 1.10 (Heine-Borel) *A subset $K \subset \mathbb{R}^n$ is compact iff it is closed and bounded.*

The non-trivial part in the proof of the Heine-Borel Theorem is the following result attributed to Weierstrass. The property that, in a given metric space, every bounded sequence has a convergent subsequence is sometimes called the *Bolzano-Weierstrass property.*

Theorem 1.11 (Bolzano-Weierstrass) *Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Proof: This proof is constructive in the sense that we will give a procedure to construct a convergent subsequence starting from an arbitrary bounded sequence. In fact, this subsequence will be a Cauchy sequence by construction, and hence, as $\mathbb{R}^n$ is complete, it converges.

Let (x_i) denote the given sequence. As (x_i) is bounded, there is $M > 0$ such that $x_i \in [-M, M]^n$, for all i. The set $[-M, M]^n$ is a cube of side $2M$. The first element of the subsequence is chosen as follows: partition C_0 into 2^n cubes of side M, and pick C_1 to be one of these smaller cubes that contains an infinite number of elements of the sequence (x_i), i.e., $x_i \in C_1$, for infinitely many i. Such a cube exists because there are only a finite number of cubes and an infinite number of elements in the sequence. Let i_1 be the smallest index such that $x_{i_1} \in C_1$. This is the first element of the subsequence.

To find the second element, first form a new sequence (y_i) by deleting from (x_i), the element x_{i_1} and all $x_j \notin C_1$. Now, repeat the procedure described in the previous paragraph but with (x_i) replaced by (y_i), and $[-M, M]^n$ replaced by C_1. This will yield a subcube C_2 of C_1 of size $M/2$, and an element y_{i_1}. Put $x_{i_2} = y_{i_1}$. This is the second element of the subsequence.

Repeat this ad infinitum to obtain a subsequence (x_{i_j}) of (x_i). By construction we will have that for every j, and every $k \geq j$, $x_{i_k} \in C_j$, where C_j is a cube of size $2M/2^j$. Therefore, (x_{i_j}) is a Cauchy sequence and hence it converges. ∎

Exercise 1.9 *Complete the proof of Theorem 1.10.*

Note that for any metric space (X, d), and any $X_0 \subset X$ we can use the same distance function d on X_0, i.e., the restriction d_0 (often simply denoted by d) of d to X_0, to make X_0 into a metric space. If X_0 is a compact subset of X, (X_0, d) is a compact metric space.

The following is a useful property.

Exercise 1.10 *Let (K, d) be a compact metric space, and G a closed subset of K. Prove that G is compact.*