Chapter 27

Limits and Topology

27.1 Introduction

This chapter opens a line of mathematical thought and methods which is quite different from purely set-theoretical, algebraic and formally logical approaches: topology and calculus. Generally speaking this perspective is about the "logic of space", which in fact explains the Greek etymology of the word "topology", which is "logos of topos", i.e., the theory of space. The "logos" is this: We learned that a classical type of logical algebras, the Boolean algebras, are exemplified by the power sets 2^a of given sets a, together with the logical operations induced by union, intersection and complementation of subsets of a (see volume 1, chapter 3). The logic which is addressed by topology is a more refined one, and it appears in the context of convergent sequences of real numbers, which we have already studied in volume 1, section 9.3, to construct important operations such as the n-th root of a positive real number. In this context, not every subset of $\mathbb{R}$ is equally interesting. One rather focuses on subsets $C \subset \mathbb{R}$ which are "closed" with respect to convergent sequences, i.e., if we are given a convergent sequence $(c_i)_i$ having all its members $c_i \in C$, then $l = \lim_{i\to\infty} c_i$ must also be an element of C. This is a useful property, since mathematical objects are often constructed through limit processes, and one wants to be sure that the limit is contained in the same set that the convergent series was initially defined in.

Actually, for many purposes, one is better off with sets complementary to closed sets, and these are called open sets. Intuitively, an open set

O in $\mathbb{R}$ is a set such that with each of its points x, a small interval of points to the left and to the right of x is still contained in O. So one may move a little around x without leaving the open set. Again, thinking about convergent sequences, if such a sequence is outside an open set, then its limit l cannot be in O since otherwise the sequence would eventually approach the limit l and then would stay in the small interval around l within O.

In the sequel, we shall not develop the general theory of topological spaces, which is of little use in our elementary context. We shall only deal with topologies on real vector spaces, and then mostly only of finite dimension. However, the axiomatic description of open and closed sets will be presented in order to give at least a hint of the general power of this conceptualization. There is also a more profound reason for letting the reader know the axioms of topology: It turns out that the open sets of a given real vector space V form a subset of the Boolean algebra 2^V which in its own right (with its own implication operator) is a Heyting algebra! Thus, topology is really a kind of spatial logic, however not a plain Boolean logic, but one which is related to intuitionistic logic. The point is that the double negation (logically speaking) of an open set is not just the complement of the complement, but may be an open set larger than the original. In other words, if it comes to convergent sequences and their limits, the logic involved here is not the classical Boolean logic. This is the deeper reason why calculus is sometimes more involved than discrete mathematics and requires very diligent reasoning with regard to the objects it produces.

27.2 Topologies on Real Vector Spaces

Throughout this section we work with the n-dimensional real vector space $\mathbb{R}^n$. The scalar product $(?,?)$ in $\mathbb{R}^n$ gives rise to the norm $\|x\| = \sqrt{(x,x)} = \sqrt{\sum_i x_i^2}$ of a vector $x = (x_1, x_2, \ldots x_n) \in \mathbb{R}^n$. Recall that for $n = 1$ the norm of x is just the absolute value of x. Actually, the theory developed here is applicable to any finite-dimensional real vector space which is equipped with a norm, and to some extent even for any infinite-dimensional real vector space with norm, but we shall only on very rare occasions encounter this generalized situation. In the following, we shall use the distance function or metric d defined through the given norm via $d(x,y) = \|x - y\|$, as defined in volume 1, section 24.3. Our first definition introduces the elementary type of sets used in the topology of real vector spaces:

Definition 175 *Given a positive real number ε, and a point $x \in \mathbb{R}^n$, the* ε-cube around x *is the set*

$$K_\varepsilon(x) = \{y \mid |y_i - x_i| < \varepsilon, \textit{ for all } i = 1, 2, \dots n\},$$

whereas the ε-ball around x *is the set*

$$B_\varepsilon(x) = \{y \mid d(x, y) < \varepsilon\}.$$

Example 98 To give a geometric intuition of the preceding concepts, consider the concrete situation for real vector spaces of dimensions 1, 2 and 3.

On the real line $\mathbb{R}$ the ε-ball and the ε-cube around x reduce to the same concept, namely the open interval of length 2ε with midpoint x, i.e., $]x - \varepsilon, x + \varepsilon[$.

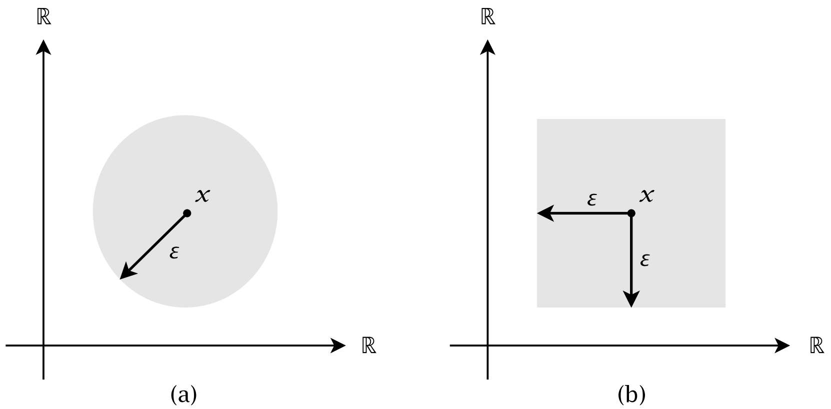

Fig. 27.1. The ε-ball (a) and ε-cube (b) around x in $\mathbb{R}^2$. The boundaries are *not* part of these sets.

On the Euclidean plane $\mathbb{R}^2$, the ε-ball around x is a disk with center x and radius ε. The boundary[1], a circle with center x and radius ε, is *not* part

[1] The precise definition of "boundary" is not needed now and will be given in definition 199.

of the disk. The ε-cube is a square with center x with distances from the center to the sides equal to ε. Again, the sides are not part of the square (figure 27.1).

The situation in the Euclidean space $\mathbb{R}^3$ explains the terminology used. In fact, the ε-ball around x is the sphere with center x and radius ε and the ε-cube is the cube with center x, where the distances from the center to the sides are equal to ε, see figure 27.2.

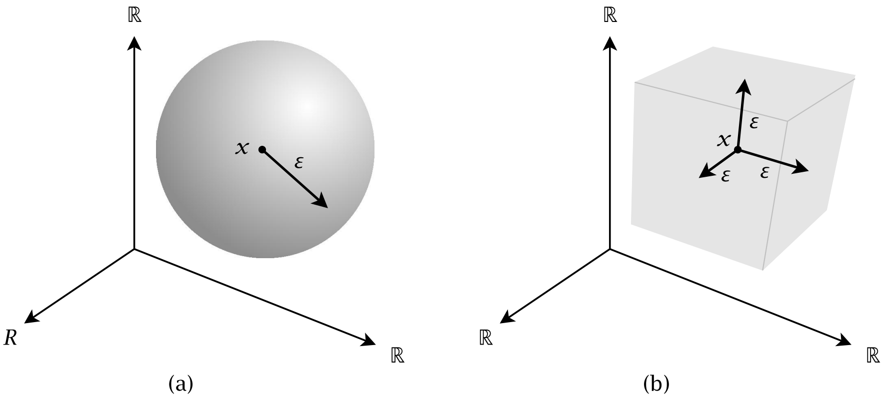

Fig. 27.2. The ε-ball (a) and ε-cube (b) around x in $\mathbb{R}^3$. The boundaries are *not* part of these sets.

The fact that both concepts, considered topologically, are in a sense equivalent, is embodied by the following lemma.

Lemma 230 *For a subset $O \subset \mathbb{R}^n$, the following properties are equivalent:*

(i) *For every $x \in O$, there is a real number $\varepsilon > 0$ such that $K_\varepsilon(x) \subset O$.*

(ii) *For every $x \in O$, there is a real number $\varepsilon > 0$ such that $B_\varepsilon(x) \subset O$.*

Proof Up to translation, it is sufficient to show that for every $\varepsilon > 0$, there is a positive real number δ such that $B_\delta(0) \subset K_\varepsilon(0)$, and conversely, there is a positive real number δ' such that $K_{\delta'}(0) \subset B_\varepsilon(0)$. For the first claim, take $\delta = \varepsilon$. Then $z = (z_1, \ldots z_n) \in B_\delta(0)$ means $\sum_i z_i^2 < \varepsilon^2$, so for every i, $|z_i| < \varepsilon$, i.e., $z \in K_\varepsilon(0)$. For the second claim, take $\delta' = \frac{\varepsilon}{\sqrt{n}}$. Then $z = (z_1, \ldots z_n) \in K_{\delta'}(0)$ means $|z_i| < \frac{\varepsilon}{\sqrt{n}}$, i.e., $\sum_i z_i^2 < n \cdot \frac{\varepsilon^2}{n}$, whence $\|z\| < \varepsilon$, i.e., $z \in B_\varepsilon(0)$. $\square$

Definition 176 *A subset* $O \subset \mathbb{R}^n$ *is called* open (in $\mathbb{R}^n$)*, iff it has the equivalent properties from definition 230. A subset* $C \subset \mathbb{R}^n$ *is called* closed (in $\mathbb{R}^n$)*, iff its complement* $\mathbb{R}^n - C$ *is open.*

Example 99 Figure 27.3 shows an open set O in $\mathbb{R}^2$ and illustrates alternative (ii) of lemma 230. Taking an arbitrary point x_1 in the open set, there is an open ball around x_1 (shown in dark gray) that is entirely contained in the open set. Two magnifications exhibit points x_2, x_3 and x_4 increasingly close to the boundary, but always an open ball can be found that lies within O, since the boundary of O is not part of O itself.

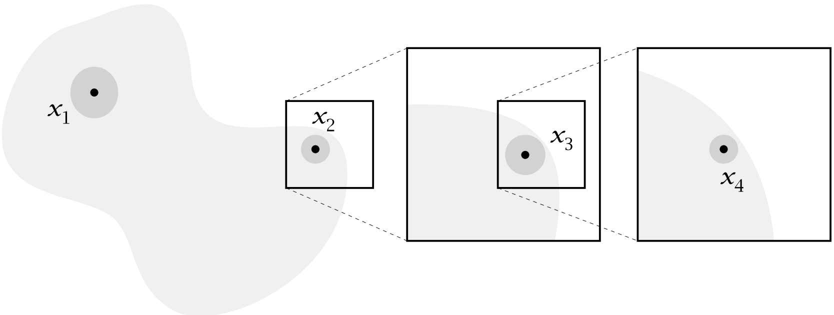

Fig. 27.3. An open set in $\mathbb{R}^2$.

In contrast, figure 27.4 shows the same set, but now it includes its boundary. Again an open ball around x_1 lies within the set, but choosing a point x_2 *on* the boundary, no ε-ball can be found that is entirely contained in the set, however small ε may be. Thus this set cannot be open. In fact, it is closed, as its complement is open.

Note that there are sets that are *both* open and closed. In $\mathbb{R}^n$ the entire set $\mathbb{R}^n$ and the empty set $\emptyset$ are both open and closed. There are also sets that are *neither* open nor closed, for example, in $\mathbb{R}$, the interval $[a, b[$ that includes a, but not b, is neither open nor closed.

Exercise 133 Show that every ball $B_\varepsilon(x)$ and every cube $K_\varepsilon(x)$ is open.

Exercise 134 Use the triangle inequality for distance functions (volume 1, proposition 213) to show that the intersection of any two balls $B_{\varepsilon_x}(x)$, $B_{\varepsilon_y}(y)$ and any two cubes $K_{\varepsilon_x}(x)$, $K_{\varepsilon_y}(y)$ is open.

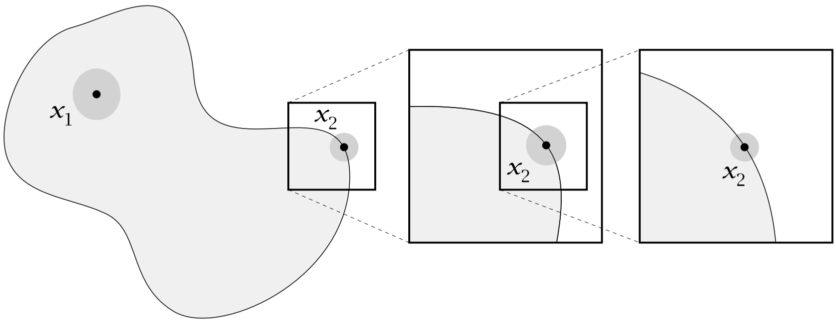

Fig. 27.4. A closed set in $\mathbb{R}^2$.

Sorite 231 *We are considering subsets of $\mathbb{R}^n$. Then:*

(i) *The empty set $\emptyset$ and the total space $\mathbb{R}^n$ are open.*

(ii) *The intersection $U \cap V$ of any two open sets U and V is open.*

(iii) *The union $\bigcup_\iota U_\iota$ of any (finite or infinite) family $(U_\iota)_\iota$ of open sets is open.*

Exercise 135 Use exercises 133 and 134 to give a proof of the properties of sorite 231.

Remark 30 More generally, a *topology* on a set X is a set $\mathcal{T}$ of subsets of X satisfying as axioms the properties of sorite 231.

Example 100 Here is a seemingly exotic, but crucial relation to logical algebras: The set *Open*$(\mathbb{R}^n)$ of open sets in $\mathbb{R}^n$ becomes a Heyting algebra by the following definitions: The maximum and minimum are $\mathbb{R}^n$ and $\emptyset$, respectively, the meet $U \wedge V$ is the intersection $U \cap V$, the join $U \vee V$ is the union $U \cup V$, and the implication $U \Rightarrow V$ is the union $\bigcup_{O \cap U \subset V} O$. (Give a proof of the Heyting properties thus defined.)

Classical two-valued logic: For any non-empty set A, consider the topology consisting of the open sets $\bot = \emptyset$ and $\top = A$. With $\vee$ and $\wedge$ as above, define $\neg U = (U \Rightarrow \bot)$. Then $\neg\top = \bigcup_{O \cap \top \subset \bot} O = \bot$ and $\neg\bot = \bigcup_{O \cap \bot \subset \top} O = \top$. These definitions satisfy the properties of a Boolean algebra.

A three-valued logic: We choose a set A, with the topology consisting of the open sets $\bot = \emptyset$, $\top = A$ and a third set X, with $X \neq \emptyset$ and $X \neq A$. Again $\neg U = (U \Rightarrow \bot)$, and we have: $\neg\top = \bot$, $\neg\bot = \top$ and $\neg X = \bot$. This last equation shows that this logic is not a Boolean algebra, since it is not the case that $x = \neg\neg x$ for all x.

A fuzzy logic: Let $A = [0, 1[$ with the topology of all intervals $I_x = [0, x[\subset A$. We have $I_x \vee I_y = I_{\max(x,y)}$ and $I_x \wedge I_y = I_{\min(x,y)}$, as well as $\bot = \emptyset$ and $\top = A$. The implication is $I_x \Rightarrow I_y = \top$, if $x \leq y$, and $I_x \Rightarrow I_y = I_y$, if $x > y$. This logic is not Boolean either.

The next definition establishes the connection to convergent sequences.

Definition 177 *A sequence $(c_i)_i$ of elements in $\mathbb{R}^n$ is called* convergent *if there is a vector $c \in \mathbb{R}^n$ such that for every $\varepsilon > 0$, there is an index N with $c_i \in B_\varepsilon(c)$ for $i > N$. Equivalently, we may require that for every $\varepsilon > 0$, there is an index M with $c_i \in K_\varepsilon(c)$ for $i > M$. If $(c_i)_i$ converges to c, one writes $\lim_{i\to\infty} c_i = c$. A sequence which does not converge is called* divergent*.*

A sequence $(c_i)_i$ of elements in $\mathbb{R}^n$ is called a Cauchy sequence*, if for every $\varepsilon > 0$, there is an index N with $c_i \in B_\varepsilon(c_j)$ for $i, j > N$. Equivalently, we may require that for every $\varepsilon > 0$, there is an index M with $c_i \in K_\varepsilon(c_j)$ for $i, j > M$.*

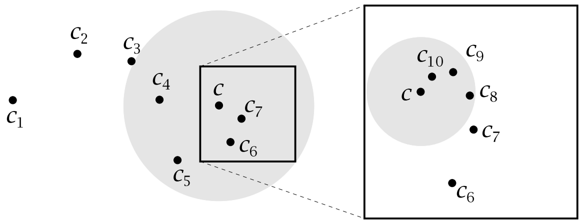

Fig. 27.5. The sequence $(c_i)_i$ converges to c. A given ε-ball around c contains all c_i for $i > 3$. In the magnification, another, smaller, ε-ball contains all c_i for $i > 7$.

Observe that this definition coincides with the already known concept of convergent and Cauchy sequences in the case $n = 1$. For example, because the ε-cube around x corresponds to the interval $]x - \varepsilon, x + \varepsilon[$ in $\mathbb{R}$, the expression $c_i \in K_\varepsilon(c_j)$ corresponds to $c_i \in\,]c_j - \varepsilon, c_j + \varepsilon[$, which in turn is equivalent to $|c_i - c_j| < \varepsilon$.

Exercise 136 Give a proof of the claimed equivalences in definition 177.

Convergence of a sequence in $\mathbb{R}^n$ is equivalent to the convergence of each of its component sequences:

Proposition 232 *For a sequence $(c_i)_i$ of elements in $\mathbb{R}^n$, and $j = 1, 2, \dots n$, we denote by $(c_{i,j})_i$ the j-th projection of $(c_i)_i$, whose i-th member $c_{i,j}$ is the j-th coordinate of the vector c_i. Then $(c_i)_i$ is convergent (Cauchy), iff all its projections $(c_{i,j})_i$ for $j = 1, 2, \dots n$ are so. Therefore, a sequence is convergent, iff it is Cauchy, and then the limit $\lim_{i\to\infty} c_i$ is uniquely determined. It is in fact the vector whose coordinates are the limits of the coordinate sequences, i.e., $(\lim_{i\to\infty} c_i)_j = \lim_{i\to\infty} c_{i,j}$.*

Proof We make use of the characterization in definition 177 of convergent or Cauchy sequences by means of cubes $K_\varepsilon(x)$. In this setting, $y \in K_\varepsilon(x)$ is equivalent to $y_j \in K_\varepsilon(x_j)$ for all projections y_j, x_j of the vectors $y = (y_1, \dots y_n), x = (x_1, \dots x_n)$ for $j = 1, \dots n$. The claims follow immediately from this fact. □

Convergent sequences provide an important characterization of closed sets:

Proposition 233 *For a subset $C \subset \mathbb{R}^n$, the following two properties are equivalent:*

(i) *The set C is closed.*

(ii) *Every Cauchy sequence $(c_i)_i$ with members $c_i \in C$ has its limit $\lim_{i\to\infty} c_i$ in C.*

Proof Suppose that C is closed and assume that the limit $c = \lim_{i\to\infty} c_i$ is in the open complement $D = \mathbb{R}^n - C$. Then there is an open ε-ball $B_\varepsilon(c) \subset D$. But there is an index N such that $i \geq N$ implies $c_i \in B_\varepsilon(c)$, a contradiction to the hypothesis that all c_i are in C. Suppose that C is not closed. Then D is not open. So there is an element $c \in D$ such that for every $i \in \mathbb{N}$, there is an element $c_i \in B_{\frac{1}{i+1}}(c) \cap C$. But then the sequence $(c_i)_i$ converges to c. □

Not every sequence is convergent, but if its members are bounded, we may extract a convergent "subsequence" from it. Boundedness is defined as follows:

Definition 178 *A* bounded sequence *is a sequence $(c_i)_i$ such that there is a real number R such that for all i, $c_i \in B_R(0)$.*

Intuitively for a bounded sequence, one can find a ball, such that the entire sequence lies within this ball, i.e., members of the sequence do not "grow indefinitely". Here is an important class of bounded sequences:

Lemma 234 *A Cauchy sequence is bounded.*

Proof This is immediate. □

Of course, the converse is false, as can be seen in the trivial example $(c_i = (-1)^i)_i$, whose members all lie in the open interval between -2 and 2. But we may extract parts of bounded sequences which are Cauchy:

Definition 179 *For a sequence $(c_i)_i$, a* subsequence $(d_i)_i$ of $(c_i)_i$ *is a sequence $(d_i)_i$ defined by an ordered injection $s : \mathbb{N} \to \mathbb{N}$, i.e., $n < m$ implies $s(n) < s(m)$, by means of $d_i = c_{s(i)}$.*

Exercise 137 Show that a subsequence $(e_i)_i$ of a subsequence $(d_i)_i$ of a sequence $(c_i)_i$ is a subsequence of $(c_i)_i$.

Proposition 235 (Bolzano-Weierstrass) *Every bounded sequence $(c_i)_i$ has a convergent subsequence.*

Proof For the proof of this theorem, we need auxiliary closed sets, namely closed cubes. A closed cube is a set of the form $K = \prod_{i=1,2,\ldots n}[a_i, b_i]$ for a sequence $a_i < b_i$ of pairs of real numbers. Such a cube K is the union of 2^n closed subcubes K^j, with $j = 1, 2, \ldots 2^n$, where each cube is defined by either the lower interval $[a_i, (a_i + b_i)/2]$ or the upper interval $[(a_i + b_i)/2, b_i]$ in the i-th coordinate. Clearly, the successive subdivision cubes $K^{j_1,j_2,\ldots j_k}$ are contained in cubes $K_\varepsilon(x)$ for any positive ε as k tends to infinity. Now, since $(c_i)_i$ is bounded, it is contained in a closed cube K. We define our convergent subsequence: Begin by taking $d_0 = c_0$. Then one of the subdivision cubes K^{j_1} contains the c_i for an infinity of indices. Take $d_1 = c_{i_1}$ with the first index $i_1 > 0$ such that $c_{i_1} \in K^{j_1}$. Then at least one of its subdivision cubes K^{j_1,j_2} contains the c_i for an infinity of indexes larger than i_1. Take the first index i_2 such that $c_{i_2} \in K^{j_1,j_2}$ and set $d_2 = c_{i_2}$. Proceeding with this procedure, we thereby define a subsequence $(d_i)_i$ of $(c_i)_i$ which is contained in progressively smaller subdivision cubes. This is a Cauchy sequence, and the proposition is proved. □

Example 101 Figure 27.6 shows a bounded sequence, where the upper and lower bounds are indicated by dashed lines. A convergent subsequence is emphasized through heavy dots.

A sequence contained in a closed set C doesn't necessarily contain any converging subsequence, an example being the sequence $(c_i = i)_i$ of natural numbers, contained in the closed set $\mathbb{R}$. But if the closed set C is bounded, i.e., if there is a radius R such that $x \in B_R(0)$ for all $x \in C$, then a fortiori, any sequence in C is bounded. But then, by the Bolzano-Weierstrass theorem, it has a convergent subsequence and its limit must

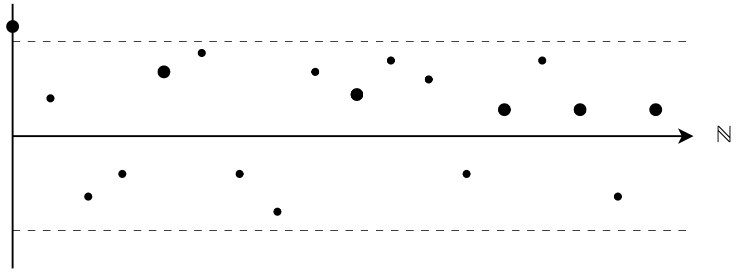

Fig. 27.6. A convergent subsequence (heavy dots) of a bounded sequence.

be an element of C by proposition 233. So every sequence in C has a convergent subsequence which converges within C! This type of closed sets is extremely important in the entire calculus and deserves its own name.

Proposition 236 *For a subset $C \subset \mathbb{R}^n$, the following properties are equivalent:*

(i) *The set C is closed and bounded.*

(ii) *Every sequence $(c_i)_i$ in C has a subsequence which converges to a point in C.*

(iii) *If $(U_i)_i$ is a (finite or infinite) family of open sets such that $C \subset \bigcup_i U_i$ (a so-called* open covering *of C), then there is a finite subfamily $U_{i_1}, \ldots U_{i_k}$ which also covers C, i.e., $C \subset \bigcup_j U_{i_j}$ (a* subcovering *of $(U_i)_i$).*

Proof (i) implies (ii): Let C be closed and bounded. A sequence $(c_i)_i$ in C has a convergent subsequence by proposition 235. Since C is closed, the limit of the subsequence is in C by proposition 233.

(ii) implies (i): If C is not bounded, then, evidently, there is a sequence $(c_i)_i$ which tends to infinity, so no subsequence can converge. If C is not closed, again by proposition 233, it contains a Cauchy sequence $(c_i)_i$ which has its limit outside C. But then every subsequence of this sequence converges to the same point outside C.

Let us now prove the equivalence of the first and third properties.

(iii) implies (i): If C is not bounded, then the open covering $(U_i = K_{i+1}(0))_i$ of $\mathbb{R}^n$ has no finite subcovering containing C. If C is bounded, but not closed, then let $x = (x_1, \ldots x_n) \notin C$ be a point such that $K_{\frac{1}{2^j}}(x) \cap C \neq \emptyset$ for all $j \in \mathbb{N}$. Take the following open covering of C. Start with the open set $U_0 = \mathbb{R}^n - \prod_i [x_i - 1, x_i + 1]$, complement of the closed cube $\prod_i [x_i - 1, x_i + 1]$. Then take the open set, $U_j = K_2(x) - \prod_i [x_i - \frac{1}{2^j}, x_i + \frac{1}{2^j}]$ for $j = 1, 2, \ldots$. This family of open sets covers $\mathbb{R}^n - \{x\}$, hence also C, but, because of the choice of x, none of its finite subcoverings contains all of C

(i) implies (iii): The converse is more delicate. Suppose that C is closed and bounded. The strategy is this: We first construct a denumerable subcovering $(U_{i_j})_{j\in\mathbb{N}}$ of C. Suppose that no finite subfamily covers C. Then for all finite subfamilies $U_{i_0}, U_{i_1}, \ldots U_{i_m}$, there is an element $c_m \in C - \bigcup_{j=0}^{m} U_{i_j}$. Since C is closed and bounded, we may even suppose that $(c_m)_m$ converges to $c \in C$. But then there is an open set $U_{i_{m_0}}$ which contains c, since $(U_{i_j})_{j\in\mathbb{N}}$ covers C. This means, by construction of $(c_m)_m$, that the members of this convergent sequence stay outside some open cube $K_\varepsilon(c)$ for $m \to \infty$, a contradiction. We now construct a denumerable subcovering of C. Clearly, if U_k is a member of our covering and if $x \in U_k$, then there is an open cube $K = \prod_i]\xi_i, \eta_i[$ which is contained in U_k, contains x, and such that its interval points ξ_i, η_i are all rational numbers. So U_k is covered by a family of open cubes with rational boundary numbers. The denumerable family $(K_r)_{r\in\mathbb{N}}$ of all these cubes, when summed up for all U_k of the given covering, also cover C, and each K_r is contained in an open set $U_{k(r)}$. Therefore the open subcovering $(U_{k(r)})_r$ is denumerable, what was claimed, and we are done. □

Definition 180 *A set $C \subset \mathbb{R}^n$ is called* compact*, iff it has the equivalent properties described in proposition 236.*

Exercise 138 Show that a compact set in $\mathbb{R}$ has a minimum and a maximum.

Proposition 237 *The Cartesian product $X \times Y \subset \mathbb{R}^{m+n}$ of two compact sets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ is compact.*

Proof First, it is clear that the Cartesian product of two bounded sets is bounded. Next, we show that the complement of $X \times Y$ is open. We have $\mathbb{R}^{m+n} - X \times Y = ((\mathbb{R}^m - X) \times \mathbb{R}^n) \cup (\mathbb{R}^m \times (\mathbb{R}^n - Y))$. We show that $(\mathbb{R}^m - X) \times \mathbb{R}^n$ is open, the other set $X \times (\mathbb{R}^n - Y)$ being then open for the same reason after exchanging left and right factors. Now, let $(x, y) \in (\mathbb{R}^m - X) \times \mathbb{R}^n$. Then there is a cube $K_\varepsilon(x) \subset \mathbb{R}^m - X$ in $\mathbb{R}^m$. Since no conditions are imposed on y, we have $(x, y) \in K_\varepsilon(x, y) \subset (\mathbb{R}^m - X) \times \mathbb{R}^n$, so $(\mathbb{R}^m - X) \times \mathbb{R}^n$ is open. □

Definition 181 *For a real number $\varepsilon > 0$ and $x \in \mathbb{R}^n$, the* closed ball *$\overline{B}_\varepsilon(x)$ is defined by $\overline{B}_\varepsilon(x) = \{y \mid d(x, y) \leq \varepsilon\}$. A* closed cube *in $\mathbb{R}^n$ is a set $[a_1, b_1] \times [a_2, b_2] \times \ldots [a_n, b_n]$ for pairs $a_i \leq b_i, i = 1, 2, \ldots n$. In particular, we have a closed cube $\overline{K}_\varepsilon(x) = [x_1 - \varepsilon, x_1 + \varepsilon] \times [x_2 - \varepsilon, x_2 + \varepsilon] \times \ldots [x_n - \varepsilon, x_n + \varepsilon]$.*

Exercise 139 Show that in $\mathbb{R}^n$ a closed cube $[a_1, b_1] \times \ldots [a_n, b_n]$ for pairs $a_i \leq b_i, i = 1, \ldots n$, as well as a closed ball $\overline{B}_\varepsilon(x)$ are compact.

Example 102 The upper half-plane $H^+ = \{(x,y) \in \mathbb{R}^2 \mid y \geq 0\}$ is a closed set in $\mathbb{R}^2$, since its complement $H^- = \{(x,y) \in \mathbb{R}^2 \mid y < 0\}$ is an open set. However H^+ is not compact, since it is not bounded (property (i)). Alternatively, we find a sequence $(c_i)_i = ((0,i))_i$ that has no convergent subsequence in H^+ (property (ii)).

The subset of integers $\mathbb{Z}$ in $\mathbb{R}$ is closed. In fact its complement in $\mathbb{R}$ is $\mathbb{Z}^C = \bigcup_{i=-\infty}^{\infty}]i, i+1[$, which is an open set, since it is the union of open intervals. $\mathbb{Z}$ is not bounded, hence not compact. Also, $U = \bigcup_{i=-\infty}^{\infty}]i - \varepsilon, i + \varepsilon[$, where $\varepsilon < 1$, is an open covering of $\mathbb{Z}$, but U contains no finite subcovering of $\mathbb{Z}$, thus property (iii) is violated.

In contrast, every closed disk $\overline{B}_r(x_0) = \{x \mid d(x_0, x) \leq r\}$, and every finite union of such closed disks, is compact.

27.3 Continuity

So far, we have only dealt with topological considerations on all of $\mathbb{R}^n$. In most practical cases, we do not have all of $\mathbb{R}^n$ at hand. For example, a function may be defined only on a closed interval of $\mathbb{R}$, or even only on an interval of type $]0,1]$, such as $f(x) = 1/x$. When applying topological considerations to such functions, we would like to deal strictly with what happens within their domains. Also, when composing two functions, the specific codomains and domains should coincide, as it is required for the composition of set functions. So we are forced to set up a minimal conceptual environment to apply topology to set functions.[2]

[2] This small extra effort will pay off: We obtain a "category" of topological spaces, i.e., topologically reasonable maps, the possibility to compose such maps and to compare topologically specified sets by means of such maps. Compare the category of matrixes, the category of sets and set maps, the category of modules and linear homomorphisms, the category of digraphs, the category of acceptors,... Later in chapter 36, we shall give a systematic account of such a conceptualization. For the moment, you just have to recognize that the present topological considerations are completely integrated within a big program of building categories of mathematical objects in order to obtain a global control of mathematical structures.

Definition 182 *Given a subset $X \subset \mathbb{R}^n$, a subset $U \subset X$ is called* open (closed) in X, *iff there is an open (closed) set $O \subset \mathbb{R}^n$ such that $U = X \cap O$. The set of open sets in X is also called the* relative topology of X. *In particular, we write $B_\varepsilon^X(x) = B_\varepsilon(x) \cap X$ and $K_\varepsilon^X(x) = K_\varepsilon(x) \cap X$ for the restrictions of the open balls and cubes, respectively, and call these open sets in X the open ball, or open cube, in X, respectively.*

Exercise 140 Show that for a given subset X of $\mathbb{R}^n$, the properties of sorite 231 are true for the open sets in X, where X plays the role of the "total space". Moreover, show that the closed sets in X are precisely the complements in X of the open sets in X.

Lemma 238 *If $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ are two subsets of Euclidean spaces, and if $f : X \to Y$ is a set map, then the following properties are equivalent.*

(i) *The inverse image $f^{-1}(U)$ of any open set U in Y is open in X.*

(ii) *For any point $x \in X$ and for any positive real number ε, there is a positive real number δ (generally depending on x and on ε) such that $f(B_\delta^X(x)) \subset B_\varepsilon^Y(f(x))$.*

(iii) *For any point $x \in X$ and for any positive real number ε, there is a positive real number δ (generally depending on x and on ε) such that $f(K_\delta^X(x)) \subset K_\varepsilon^Y(f(x))$.*

(iv) *The inverse image $f^{-1}(U)$ of any closed set U in Y is closed in X.*

(v) *For any point $x \in X$ and for any convergent sequence $(c_i)_i$ with $\lim_{i\to\infty} c_i = x$, the image sequence $(f(c_i))_i$ converges to $f(x)$.*

Proof (i) implies (ii): Since $f^{-1}(B_\varepsilon^Y(f(x)))$ is open and contains x, there is an open ball $B_\delta^X(x) \subset f^{-1}(B_\varepsilon^Y(f(x)))$. Therefore $f(B_\delta^X(x)) \subset B_\varepsilon^Y(f(x))$.

(ii) implies (i): Since every open set U is the union of open balls, its inverse image is the union of inverse images of open balls. But by (ii), the inverse image of an open ball is a union of open balls, and, therefore, open, whence (i).

The same argument yields the equivalence of (i) and (iii).

The equivalence of (i) and (iv) results from the set-theoretic fact that complements and inverse images commute.

(ii) implies (v): Let $\varepsilon > 0$. Then there is $\delta > 0$ such that $f(B_\delta^X(x)) \subset B_\varepsilon^Y(f(x))$. So by the convergence of $(c_i)_i$, there is a natural number N such that $i \geq N$ implies $c_i \in B_\delta^X(x)$. Hence $i \geq N$ implies $f(c_i) \in B_\varepsilon^Y(f(x))$, therefore $(f(c_i))_i$ converges to $f(x)$.

(v) implies (ii): Suppose (ii) is false for an $x \in X$. Then there is $\varepsilon_0 > 0$ such that for every $i \in \mathbb{N}$, there is $c_i \in B_{\frac{1}{i+1}}(x)$ with $f(c_i) \notin B_{\varepsilon_0}(f(x))$. But the sequence

$(c_i)_i$ evidently converges to x, while the images $f(c_i)$ stay outside the open ball $B_{\varepsilon_0}(f(x))$, which contradicts (v). □

Definition 183 *If $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ are two subsets of Euclidean spaces, a set map $f : X \to Y$ with the equivalent properties of lemma 238 is called* continuous. *A continuous bijection f such that its inverse f^{-1} is also continuous, is called a* homeomorphism. *The set of continuous maps $f : X \to Y$ is denoted by $Top(X, Y)$. In particular, if $Y = \mathbb{R}$, one writes $C^0(X) = Top(X, \mathbb{R})$. If in lemma 238, the conditions (ii) to (iv) are valid for a specific point x only, f is called* continuous in x. *This means that f is continuous, iff it is continuous in every x of its domain.*

Example 103 To illustrate property (i) of lemma 238, it is best to show a case where the property fails. In figure 27.7, the function f is non-continuous, as is clear by the jump at the argument x. The value at x, $f(x)$, is indicated by the heavy dot. Now, the inverse image $f^{-1}(U)$ of the open interval U, is not open, but a half-open interval, i.e., open at the left and closed at the right with x as the endpoint.

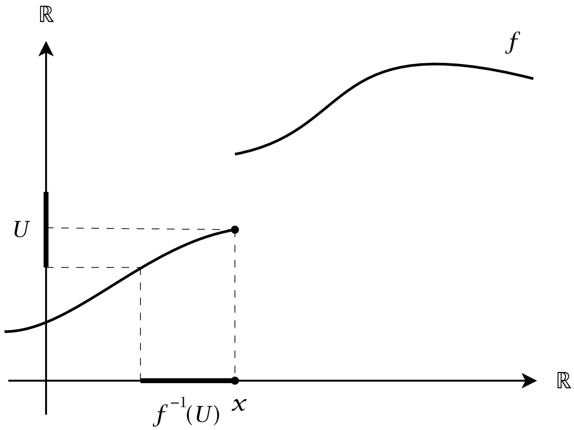

Fig. 27.7. The function f being *non*-continuous, the inverse image of the open set U is not open, in fact it is a half-open interval.

Sorite 239 *Let $X \subset \mathbb{R}^m, Y \subset \mathbb{R}^n, V \subset \mathbb{R}^s, W \subset \mathbb{R}^t, Z \subset \mathbb{R}^l$ be subsets of Euclidean spaces.*

(i) *The identity Id_X is always a homeomorphism.*

(ii) *If $f : X \to Y$ and $g : Y \to Z$ are continuous, then their composition $g \circ f$ is also continuous.*

(iii) *If $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$ are continuous, then $(h \circ g) \circ f = h \circ (g \circ f) = h \circ g \circ f$.*

(iv) *If $f : X \to Y$ and $u : U \to V$ are continuous, then so is the Cartesian product map $f \times u : X \times U \to Y \times V$.*

(v) *The projections $pr_X : X \times Y \to X$ and $pr_Y : X \times Y \to Y$ are continuous.*

(vi) *If $a : U \to X$ and $b : U \to Y$ are continuous maps, then so is the universal map $(a, b) : U \to X \times Y$ associated with a and b (see volume 1, proposition 57).*

Proof (i) is evident.

(ii) follows from the fact that for an open set $U \subset Z$, we have $(g \circ f)^{-1}(U) = g^{-1}(f^{-1}(U))$, and since $V = f^{-1}(U)$ is an open set so is $g^{-1}(V)$.

(iii) Associativity is clear, since it is true for any set maps.

(iv) It suffices to show that the inverse image of an open cube $K_\varepsilon^{X \times Y}(x, y)$ under $f \times u$ is open. But we have $K_\varepsilon^{X \times Y}(x, y) = K_\varepsilon^X(x) \times K_\varepsilon^Y(y)$, therefore $(f \times u)^{-1}(K_\varepsilon^{X \times Y}(x, y)) = f^{-1}(K_\varepsilon^X(x)) \times u^{-1}(K_\varepsilon^Y(y))$, and this is open.

For (v), observe that the cube $K_\varepsilon^{X \times Y}(x, y)$ is mapped by pr_X into the cube $K_\varepsilon^X(x)$, since cube elements are characterized coordinatewise. Similarly for the second projection.

As to (vi), if at a point $v \in U$, we have $a(K_\delta(v)) \subset K_\varepsilon(a(v))$ and $b(K_\delta(v)) \subset K_\varepsilon(b(v))$, then $(a, b)(K_\delta(v)) \subset K_\varepsilon^X(a(v)) \times K_\varepsilon^Y(b(v)) = K_\varepsilon^{X \times Y}((a, b)(v))$. □

These seemingly innocent general properties of continuous maps have a large number of very important consequences concerning the continuity of functions which are known from the theory of polynomials and from linear geometry. The crucial fact is this:

Lemma 240 *The maps of addition, $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and multiplication, $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, are continuous. The inversion $?^{-1} : \mathbb{R}^* \to \mathbb{R}^*$ is continuous.*

Proof By the basic properties of the real number arithmetic, we have, for the addition, $+(K_{\varepsilon/2}(x, y)) \subset K_\varepsilon(x + y)$.

For the product $x \cdot y$, we have $|(x + v) \cdot (y + \mu) - x \cdot y| \le |x||\mu| + |y||v| + |\mu||v|$. If $xy \ne 0$, take $\delta = \min\{\varepsilon/3|x|, \varepsilon/3|y|, \sqrt{\varepsilon/3}\}$. If $x = 0$ and $y \ne 0$, then take $\delta = \min\{\varepsilon/2|y|, \sqrt{\varepsilon/2}\}$. If $x \ne 0$ and $y = 0$, then take $\delta = \min\{\varepsilon/2|x|, \sqrt{\varepsilon/2}\}$. If $x = y = 0$, take $\delta = \sqrt{\varepsilon}$. We then obtain $(K_\delta(x, y)) \subset K_\varepsilon(x \cdot y)$.

The third statement is left as an exercise for the reader. □

Proposition 241 *A polynomial function $P : \mathbb{R}^n \to \mathbb{R}$ defined by a polynomial $P \in \mathbb{R}[X_1, \ldots X_n]$ of n variables $X_1, \ldots X_n$ is continuous.*

Proof If $P = a \in \mathbb{R}$ is a constant, the polynomial function $P : \mathbb{R}^n \to \mathbb{R}$ is constant, and this is evidently continuous. In general, if $f_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots k$, are continuous, then, by proposition 240, their sum $\sum_i f_i : \mathbb{R}^n \to \mathbb{R} : x \mapsto \sum_i f_i(x)$ and their product $\prod_i f_i : \mathbb{R}^n \to \mathbb{R} : x \mapsto \prod_i f_i(x)$ are continuous since we have $\sum_i f_i = \sum \circ (f_1, \ldots f_k)$ and $\prod_i f_i = \prod \circ (f_1, \ldots f_k)$, where $(f_1, \ldots f_k) : \mathbb{R}^n \to \mathbb{R}^k$ is the universal map of Cartesian products, and where $\sum$ and $\prod$ are the continuous k-fold sum and product maps. But the polynomial function P is the sum of its monomials, so it is continuous if the monomials are so. Further, each monomial $aX^{n_1} \ldots X^{n_t}$ in P is a product of the constant a and the projection functions X_j, which are all continuous by sorite 239, hence P is continuous. □

Exercise 141 Give a proof of proposition 241 for the polynomial $P = 2X_1^2 - X_2 \cdot X_3 + 1.5$ using sorite 239 and lemma 240.

Lemma 242 *The maps of addition and multiplication, $+ : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $\cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ are continuous for the complex numbers, where we interpret $\mathbb{C}$ as the real vector space $\mathbb{R}^2$ to define its topology. Conjugation $\overline{?} : \mathbb{C} \to \mathbb{C}$ of complex numbers is a homeomorphism.*

Proof This follows immediately since these operations, when rewritten in real coordinates, are polynomial functions. So proposition 241 applies. □

Using the above general facts from sorite 239, we deduce the following theorem about continuity of matrix operations. This requires that matrixes $M \in \mathbb{M}_{m,n}(\mathbb{R})$ are viewed as vectors in some Euclidean space. We do this in the usual way by the well-known identification of $\mathbb{M}_{m,n}(\mathbb{R})$ with $\mathbb{R}^{mn}$, the Euclidean structure on $\mathbb{M}_{m,n}(\mathbb{R})$ being induced from the Euclidean structure on $\mathbb{R}^{mn}$. For example, the norm of a matrix $M = (M_{i,j})$ is $\|M\| = \sqrt{\sum_{i,j} M_{i,j}^2}$.

Proposition 243 *The following maps are all continuous:*

(i) *Addition* $+ : \mathbb{M}_{m,n}(\mathbb{R}) \times \mathbb{M}_{m,n}(\mathbb{R}) \to \mathbb{M}_{m,n}(\mathbb{R}) : (M, N) \mapsto M + N$,

(ii) *Multiplication* $\cdot : \mathbb{M}_{l,m}(\mathbb{R}) \times \mathbb{M}_{m,n}(\mathbb{R}) \to \mathbb{M}_{l,n}(\mathbb{R}) : (M, N) \mapsto M \cdot N$,

(iii) *Scalar product* $(?, ?) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

(iv) *Scalar multiplication* $\mathbb{R} \times \mathbb{M}_{m,n}(\mathbb{R}) \to \mathbb{M}_{m,n}(\mathbb{R}) : (t, M) \mapsto t \cdot M$,

(v) *Determinant function* $\det : \mathbb{M}_{n,n}(\mathbb{R}) \to \mathbb{R}$,

(vi) *Matrix transposition* $\tau : \mathbb{M}_{m,n}(\mathbb{R}) \to \mathbb{M}_{n,m}(\mathbb{R})$,

(vii) *Matrix adjunction* $Ad : \mathbb{M}_{n,n}(\mathbb{R}) \to \mathbb{M}_{n,n}(\mathbb{R})$.

Proof All the claims of this proposition are immediate from the polynomial character of the involved functions and their combinations, following sorite 239 and proposition 241. We leave the details to the reader as a useful exercise. □

A central fact about continuous maps is

Proposition 244 *The image $f(X) \subset \mathbb{R}^m$ of a compact set $X \subset \mathbb{R}^n$ under a continuous map $f : X \to \mathbb{R}^m$ is compact.*

Proof Let $(U_i)_i$ be an open covering of $f(X)$. Then the inverse image of $f(X) = \bigcup_i U_i$ is $X = \bigcup_i f^{-1}(U_i)$, an open covering of X. So there is a finite subcovering $\bigcup_{j=1}^{J} f^{-1}(U_{i_j})$ of X. Therefore $f(X) = f(\bigcup_{j=1}^{J} f^{-1}(U_{i_j})) = \bigcup_{j=1}^{J} f(f^{-1}(U_{i_j})) \subset \bigcup_{j=1}^{J} U_{i_j} \subset f(X)$, so we obtain the finite subcovering $\bigcup_{j=1}^{J} U_{i_j}$ of $f(X)$. □

In particular, by exercise 138, if we are given a continuous function $f : X \to \mathbb{R}$ on a compact set X, there are two arguments $x, y \in X$ such that $f(x) \leq f(z) \leq f(y)$ for all $z \in X$, i.e., the minimum and maximum of $f(X)$ are obtained as function values. But we do not know whether all intermediate values are obtained. This property is guaranteed by the famous *intermediate value theorem (Zwischenwertsatz)* first proved by the German mathematician Bernhard Bolzano in 1817.

Proposition 245 (Bolzano) *If $K = [a_1, b_1] \times [a_2, b_2] \times \ldots [a_n, b_n]$ for pairs $a_i \leq b_i, i = 1, 2, \ldots n$ is a closed cube in $\mathbb{R}^n$, and if $f : K \to \mathbb{R}$ is continuous, then Im(f) is a closed interval $[a, b]$, i.e., for each value c between the minimum $a = f(x)$ and maximum $b = f(y)$ of Im(f), there is an argument $z \in K$ such that $c = f(z)$.*

Proof By proposition 244, $f(K)$ is compact, i.e., closed and bounded by proposition 236. Therefore $b = \sup(f(K))$ is finite. But taking a sequence $(c_i)_i$ in $f(K)$ which converges to b, closedness of $f(K)$ implies that $b \in f(K)$. A similar argument works with $a = \inf(f(K))$, the infimum[3] of $f(K)$. Therefore there is a maximal and a minimal value for $f(K)$. Now, let $r = f(u) < s = f(v)$ for $u, v \in K$ be any two values in $f(K)$. We claim that any value $c \in [r, s]$ is taken by an argument $z \in K$, i.e., $c = f(z)$. Consider the map $\gamma : [0, 1] \to K : \xi \mapsto \xi \cdot u + (1 - \xi) \cdot v$. This is evidently a continuous map, since it is even affine. The composition $g = f \cdot \gamma : [0, 1] \to \mathbb{R}$ is continuous and we have $g(0) = r < g(1) = s$. So we have reduced the problem to a one-dimensional cube $[0, 1]$. Suppose that there is $x \in T = [r, s] - g([0, 1])$. Take the supremum y of T. Since $s \notin T$,

[3] The infimum of a non-empty set $A \subset \mathbb{R}$, which is bounded from below, i.e., there is $l < x$, for all $x \in A$, is the number $\inf(A) = -\sup(-A)$, where $-A = \{-x \mid x \in A\}$.

the supremum is an element of the interval $[r, s]$, and smaller than s. Take a sequence $(c_i)_i$, $c_i \in]y, s]$ which converges to y. By construction of y, there is a sequence $(d_i)_i$ in $[0, 1]$ with $f(d_i) = c_i$, for all i. But $[0, 1]$ is compact, so there is even a convergent subsequence $(e_i)_i$ of $(d_i)_i$, converging to $e \in [0, 1]$, say. But then, by continuity, $f(e) = f(\lim_{i\to\infty} e_i) = \lim_{i\to\infty} f(e_i) = y$, a contradiction. □

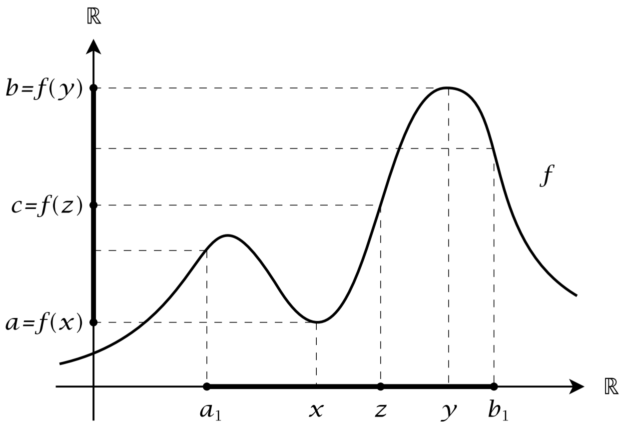

Fig. 27.8. Intermediate value theorem.

Recall that we used this result to prove proposition 219 in chapter 25.3 of volume 1.

Corollary 246 *For a polynomial $P \in \mathbb{R}[X]$ of odd degree, there is an argument $x \in \mathbb{R}$ such that $P(x) = 0$.*

Proof Since P is continuous, it suffices to find arguments $a, b \in \mathbb{R}$ such that $P(a) < 0$ and $P(b) > 0$. Let $P(x) = a_{2n+1}x^{2n+1} + a_{2n}x^{2n} + \dots a_0$. We may evidently suppose $a_{2n+1} = 1$, since the general case follows immediately from this special case. For $x \neq 0$, we write $P(x) = x^{2n+1}(1 + \frac{a_{2n}}{x} + \dots \frac{a_0}{x^{2n+1}})$. Consider positive natural numbers $x = i$ as arguments of P. If $i \to \infty$, then the summands $\frac{a_{2n}}{i} \dots \frac{a_0}{i^{2n+1}}$ converge to 0. Therefore the factor $1 + \frac{a_{2n}}{i} + \dots \frac{a_0}{i^{2n+1}}$ converges to 1. This implies that the product $P(i) = i^{2n+1}(1 + \frac{a_{2n}}{i} + \dots \frac{a_0}{i^{2n+1}})$ tends to ∞ as $i \to \infty$. For integers $i < 0$, if $i \to -\infty$, then $P(i) = i^{2n+1}(1 + \frac{a_{2n}}{i} + \dots \frac{a_0}{i^{2n+1}})$ tends to $i^{2n+1} < 0$, so we have positive and negative values and then, by proposition 245, there is an x such that $P(x) = 0$. □

This last result was used in chapter 25.1 of volume 1.

27.4 Series

This section introduces a more systematic study of sequences, and in particular sequences deduced from partial sums of given sequences. These series play a central role in the construction of basic continuous functions, but also, as we shall see later in this book under the title of Taylor series, in the reconstruction of quite general functions in terms of convergent sequences of polynomial functions.

To begin with, consider the real vector space $Sequ(\mathbb{R}, n) = (\mathbb{R}^n)^{\mathbb{N}}$ of sequences $(c_i)_i$ with values in $\mathbb{R}^n$. Recall that the sum and scalar multiplication are defined coordinatewise, i.e., $(c_i)_i + (d_i)_i = (c_i + d_i)_i$, and $\lambda(c_i)_i = (\lambda c_i)_i$ for $\lambda \in \mathbb{R}$. Denote by $C(\mathbb{R}, n)$ the subset of Cauchy or, equivalently, convergent sequences in $Sequ(\mathbb{R}, n)$.

Lemma 247 *The set $C(\mathbb{R}, n)$ is a vector subspace of $Sequ(\mathbb{R}, n)$. The map $\lim_{i\to\infty} : C(\mathbb{R}, n) \to \mathbb{R}^n : (c_i)_i \mapsto \lim_{i\to\infty} c_i$ is linear and its kernel is the sub-vector space $\mathcal{O}(\mathbb{R}, n)$ of zero sequences.*

Exercise 142 Give a proof of lemma 247. Check in particular that the statement of linearity of the map $\lim_{i\to\infty}$ is equivalent to the fact that limits of sums of sequences are the sums of their limits, whereas the product of a constant λ with the members of a convergent sequence converges to the scaling of the sequence limit by λ.

Definition 184 *Consider the following two linear endomorphisms Σ and Δ of $Sequ(\mathbb{R}, n)$:*

$$\Sigma : Sequ(\mathbb{R}, n) \to Sequ(\mathbb{R}, n) : (c_i)_i \mapsto \Sigma(c_i)_i,$$

the i-th member of $\Sigma(c_i)_i$ being $\Sigma(c_i) = \Sigma_{j=0}^{i} c_j$. The image sequence $\Sigma(c_i)_i$ is called the (associated) series of $(c_i)_i$. *And*

$$\Delta : Sequ(\mathbb{R}, n) \to Sequ(\mathbb{R}, n) : (c_i)_i \mapsto \Delta(c_i)_i,$$

the i-th member of $\Delta(c_i)_i$ being $\Delta(c_i) = c_i - c_{i-1}$ for positive i and $\Delta(c_0) = c_0$. The image $\Delta(c_i)_i$ is called the (associated) difference of $(c_i)_i$.

Lemma 248 *The endomorphisms Σ and Δ are automorphisms and inverses of each other, i.e.,*

$$\Delta \circ \Sigma = \Sigma \circ \Delta = Id_{Sequ(\mathbb{R}, n)}.$$

Proof This is immediate by a straightforward calculation, which we leave to the reader. □

This means that the inverse image $\Delta^{-1}C(\mathbb{R}, n)$ is the vector space of sequences having convergent series. If a series $\Sigma(c_i)_i$ converges, we write $\Sigma_{i=0}^{\infty} c_i$ for its limit. By the identification of Cauchy and convergent sequences, we have:

Proposition 249 *If a series $\Sigma(c_i)_i$ converges, then $(c_i)_i$ is a zero sequence, i.e,*

$$\Delta(C(\mathbb{R}, n)) \subset \mathcal{O}(\mathbb{R}, n).$$

Proof This follows from the Cauchy condition $|\Sigma(c_{i+1}) - \Sigma(c_i)| < \varepsilon$ for subsequent partial sums of sufficiently high index i of the series. □

Example 104 Given a real number q, the sequence $(q^i)_i$ gives rise to the *geometric series* $\Sigma(q^i)_i$ with general member

$$\Sigma(q^i) = 1 + q + q^2 + \ldots q^i.$$

For $q \neq 1$, one has the formula $1 + q + q^2 + \ldots q^i = \frac{1-q^{i+1}}{1-q}$. Since we have $\Sigma(q^i) = \frac{1}{1-q} - \frac{q^{i+1}}{1-q}$, convergence is a linear map, and the second summand converges to zero for $|q| < 1$, we have the very important formula

$$\Sigma_{i=0}^{\infty} q^i = \frac{1}{1-q}$$

for $|q| < 1$. Try to understand this result geometrically for the intuitive special value $q = \frac{1}{2}$.

But there are zero sequences without converging associated series:

Example 105 The *harmonic series* $\Sigma\left(\frac{1}{i+1}\right)_i$, with partial sums $\Sigma\left(\frac{1}{i+1}\right) = 1 + \frac{1}{2} + \ldots \frac{1}{i+1}$, is divergent. Nonetheless, the very similar *alternating series* $\Sigma\left(\frac{(-1)^i}{i+1}\right)_i$ is convergent. This is a special case of the following *Leibniz criterion.*

Proposition 250 *If $(c_i)_i \in Sequ(\mathbb{R}, 1)$ is a zero sequence which is monotonously decreasing, i.e., $c_i \geq c_{i+1}$ for all i, then the alternating series $\Sigma((-1)^i c_i)_i$ converges.*

Proof We are given a series with $c_0 \geq c_1 \geq \ldots$ which converges to 0. Let us show by induction on N that the partial sums $S_N = \sum_{i=0}^{N}(-1)^i c_i$ satisfy $0 \leq S_N \leq c_0$. This is true for $N = 0, 1, 2$ by immediate check. In general, if N is even, we have $S_N = S_{N-2} - c_{N-1} + c_N$, whence $S_N \leq S_{N-2} \leq c_0$, but also $S_N = S_{N-1} + c_N \geq S_{N-1} \geq 0$. If N is odd, then $S_N = S_{N-2} + c_{N-1} - c_N$, whence $S_N \geq S_{N-2} \geq 0$, but also $S_N = S_{N-1} - c_N \leq S_{N-1} \leq c_0$. Now, Cauchy's criterion for convergence requires $|S_N - S_M| < \varepsilon$ for N, M sufficiently large. But $S_N - S_M$ is just a partial sum of such an alternating series starting from $m = \min(M, N)$. If this minimum is sufficiently large, by the above, the difference is limited by c_m, which converges to 0, so we are done. □

A partially converse criterion is the famous *criterion of absolute convergence.*

Definition 185 *A series $\Sigma(c_i)_i \in Sequ(\mathbb{R}, n)$ is said to be* absolutely convergent *if the series $\Sigma(\|c_i\|)_i \in Sequ(\mathbb{R}, 1)$ converges.*

Proposition 251 *An absolutely convergent series $\Sigma(c_i)_i \in Sequ(\mathbb{R}, n)$ is convergent.*

Proof Let $\Sigma(c_i)_i$ be absolutely convergent. Then for two indexes $N \leq M$, the triangle inequality in $\mathbb{R}^n$ yields $\|\Sigma(c_M) - \Sigma(c_N)\| = \|\sum_{i=N+1}^{M} c_i\| \leq \sum_{i=N+1}^{M} \|c_i\|$, and the latter is smaller than any positive ε for M, N sufficiently large by the absolute convergence hypothesis. Therefore the Cauchy criterion yields convergence of the series. □

The next criterion gives us a large variety of absolutely convergent series at hand:

Proposition 252 *If a series $\Sigma(c_i)_i \in Sequ(\mathbb{R}, n)$ is based on a sequence $(c_i)_i$ with non-zero members such that there is a real number $0 < q < 1$ with this property: There is a natural N such that $\frac{\|c_{i+1}\|}{\|c_i\|} \leq q$ for all $i > N$, then $\Sigma(c_i)_i$ is absolutely convergent.*

Proof Since the initial portion of a sequence is irrelevant for its convergence, we may suppose that $\frac{\|c_{i+1}\|}{\|c_i\|} \leq q$ for all $i \geq 0$. Then we have $\|c_i\| \leq q^i \|c_0\|$, all $i \in \mathbb{N}$. Therefore $\Sigma(\|c_i\|) \leq \|c_0\| \cdot (1 + q + q^2 + \ldots q^i)$ which is a convergent geometric series. □

For the next result, we again interpret complex numbers as vectors in $\mathbb{R}^2$ and accordingly consider sequences with members in $\mathbb{C}$ as series in the Euclidean space $\mathbb{R}^2$.

Corollary 253 *Given a complex number $z \in \mathbb{C}$, the* power series *(involving powers of z)* $\Sigma \left(\frac{z^k}{k!}\right)_k$ *is absolutely convergent. We therefore define the complex-valued function*

$$\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

which is called the exponential function.

Proof The absolute convergence follows immediately from the ratio

$$\frac{\frac{z^{k+1}}{(k+1)!}}{\frac{z^k}{(k)!}} = \frac{z}{k+1},$$

which tends to 0 for $k \to \infty$, and the proposition 252 applies. □

27.4.1 Fundamental Properties of the Exponential Function

In this subsection, we want to deal with some technical aspects which are of general interest, but which are also crucial for the establishment of fundamental properties of the exponential function. In particular, we want to calculate the value $\exp(w+z)$, and since this involves the powers $(w+z)^k$ as functions of w and z, we need to calculate polynomials $(X+Y)^k \in \mathbb{Z}[X,Y]$ first. To this end, we need a formula for the coefficients of such polynomials. These coefficients will also play an important role in the calculus of probability, to name but one example. They are in fact omnipresent in mathematics as soon as it comes to the calculation of any combinatorial quantities.

Definition 186 *Let* $0 \le k \le n$ *be natural numbers. Then one sets*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}$$

(with the special value $0! = 1$*) and calls this rational number the* binomial coefficient n over k*.*

Here is the basic result which allows the inductive calculation of binomial coefficients:

Lemma 254 *For natural numbers* $0 \le k < n$*, we have*

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

In particular, by induction on n*, and observing that* $\binom{n}{0} = 1$*, it follows that binomial coefficients are integers.*

Proof We have

$$\begin{aligned}\binom{n}{k} + \binom{n}{k+1} &= \frac{n \cdot (n-1) \cdot \ldots (n-k+1) \cdot (k+1)}{k!(k+1)} + \\ &\quad \frac{n \cdot (n-1) \cdot \ldots (n-k+1) \cdot (n-k)}{k!(k+1)} \\ &= \frac{n \cdot (n-1) \cdot \ldots (n-k+1)}{(k+1)!}((k+1) + (n-k)) \\ &= \binom{n+1}{k+1}.\end{aligned}$$

□

The *Pascal triangle* (figure 27.9) is a graphical representation of the above result: We represent the binomial coefficients for a given n on a row and develop the coefficients from $n = 0$ on downwards. Observe the vertical symmetry axis in the triangle, which stems from the obvious fact that $\binom{n}{k} = \binom{n}{n-k}$.

$$\begin{array}{ccccccccccccccc} & & & & & & & 1 & & & & & & & \\ & & & & & & 1 & & 1 & & & & & & \\ & & & & & 1 & & 2 & & 1 & & & & & \\ & & & & 1 & & 3 & & 3 & & 1 & & & & \\ & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\ & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\ & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\ 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 \\ & & & & & & & \ldots & & & & & & & \end{array}$$

Fig. 27.9. The Pascal triangle.

This yields the coefficients of $(X+Y)^n$ as follows:

Proposition 255 *If* $n \in \mathbb{N}$*, then the polynomial* $(X+Y)^n \in \mathbb{Z}[X,Y]$ *has this representation in terms of monomials:*

$$(X+Y)^n = \sum_{k=0}^{n} \binom{n}{k} X^{n-k} \cdot Y^k = X^n + nX^{n-1} \cdot Y + \ldots nX \cdot Y^{n-1} + Y^n.$$

Proof One proves the proposition by induction on n using the recursive formula from lemma 254. This is just an exercise in reindexing sums, we therefore omit it and refer to [14]. □

This allows us to regard the expression $\exp(w+z)$ as a series of the following products:

$$\exp(w+z) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{n!}\binom{n}{k}w^{n-k}z^k = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{(n-k)!}w^{n-k}\frac{1}{k!}z^k.$$

So we are confronted with the problem of whether a product of series is the series of the products of their summands. This is precisely what the following proposition guarantees:

Proposition 256 *Identifying $C(\mathbb{C},1)$ with the vector space $C(\mathbb{R},2)$ over the Euclidean space $\mathbb{R}^2$, if $\Sigma(c_i)_i$ and $\Sigma(d_i)_i$ are absolutely convergent series in $C(\mathbb{C},1)$, then we have the Cauchy product formula*

$$\left(\sum_{i=0}^{\infty} c_i\right)\cdot\left(\sum_{i=0}^{\infty} d_i\right) = \sum_{i=0}^{\infty}\sum_{k=0}^{i} c_{i-k}d_k.$$

This is a special case of a formula guaranteeing that a series is absolutely convergent, iff it is "unconditionally" convergent, which means that it converges to the same limit for any permutation of the summation. We cannot delve into those details and refer to [14].

Proposition 256 implies the following result.

Proposition 257 *The map*

$$\exp : \mathbb{C} \to \mathbb{C}^*$$

is a surjective continuous group homomorphism from the additive group of complex numbers to the multiplicative group of non-zero complex numbers, i.e., $\exp(0) = 1$ *and* $\exp(w+z) = \exp(w)\cdot\exp(z)$ *for all* $w, z, \in \mathbb{C}$. *There is a number* $\pi = 3.1415926\ldots$ *such that*

$$Ker(\exp) = i2\pi\mathbb{Z}.$$

In particular, $\mathbb{C}/i2\pi\mathbb{Z} \xrightarrow{\sim} \mathbb{C}^*$. *The inverse image of the unit circle subgroup* $\boldsymbol{U} \subset \mathbb{C}^*$ *is the additive group* $i\cdot\mathbb{R}$, *in particular,*

$$\boldsymbol{U} \xrightarrow{\sim} i\cdot\mathbb{R}/i2\pi\mathbb{Z} \xrightarrow{\sim} \mathbb{R}/\mathbb{Z}.$$

This combines to the group isomorphism $\mathbb{R} \times \boldsymbol{U} \xrightarrow{\sim} \mathbb{C}^* : (r, u) \mapsto \exp(r) \cdot u$*, which is called the* polar coordinate representation *of (non-zero) complex numbers. The uniquely determined angle* $-\pi < \theta \leq \pi$*, such that* $u = \exp(i \cdot \theta) \in \boldsymbol{U}$ *in the polar coordinate representation* $z = \exp(r) \cdot u$ *is denoted by* $\arg(z)$*. The Euler formula* $\exp(i \cdot \theta) = \cos(\theta) + i \sin(\theta)$ *established in proposition 210 in volume 1 (we used the symbol* $A(\theta)$ *for* $\exp(i \cdot \theta)$ *there), implies the representation of the sine and cosine functions, which are both continuous, in terms of power series:*

$$\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots (-1)^n \frac{\theta^{2n}}{(2n)!} + \cdots$$
$$\sin(\theta) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} \cdots (-1)^n \frac{\theta^{2n+1}}{(2n+1)!} + \cdots$$

The Euler formula implies this alternative definition of the sine and cosine functions:

$$\cos(\theta) = \frac{1}{2}(\exp(i \cdot \theta) + \exp(-i \cdot \theta)),$$
$$\sin(\theta) = \frac{1}{2i}(\exp(i \cdot \theta) - \exp(-i \cdot \theta)).$$

The restriction $\exp|_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}_+$ *is continuous and ordered*[4] *isomorphism of the additive group of* $\mathbb{R}$ *onto the multiplicative group* $\mathbb{R}_+$ *of positive real numbers. Its inverse* $\log : \mathbb{R}_+ \to \mathbb{R}$ *is called the* (natural) logarithm*. In particular,* $\log(1) = 0$*, and* $\log(x \cdot y) = \log(x) + \log(y)$ *for all* $x, y \in \mathbb{R}_+$*.*

The number $e = \exp(1) = \sum_{k=0}^{\infty} \frac{1}{k!} = 2.7182818\ldots$ *is called the* Euler number*; it is also equal to* $\lim_{n\to\infty}(1 + \frac{1}{n})^n$*. For a rational number* $\frac{p}{q}$ *with* $q > 0$*, we have* $\exp(\frac{p}{q}) = e^{\frac{p}{q}} = (\sqrt[q]{e})^p$*. The general value* $\exp(z)$ *for* $z \in \mathbb{C}$ *is therefore also written as* e^z*.*

Proof By proposition 256, exp is a group homomorphism, i.e., for all $w, z \in \mathbb{C}$, $\exp(w + z) = \exp(w) \cdot \exp(z)$. In particular, $1 = \exp(0) = \exp(w + (-w)) = \exp(w) \cdot \exp(-w)$, whence $\exp : \mathbb{C} \to \mathbb{C}^*$ is a group homomorphism into the multiplicative group of non-zero complex numbers. Moreover, exp is continuous. In fact, for any $w \in \mathbb{C}$, we have $\exp(w + z) - \exp(w) = \exp(z)(\exp(w) - 1)$. So we have to show that $\exp(w) - 1 \to 0$ if $w \to 0$. But $\|\exp(w) - 1\| \leq \|w\| \cdot \sum_k \frac{\|w\|^k}{(k+1)!} \leq \|w\| \cdot \sum_k \frac{\|w\|^k}{k!} \leq \|w\| \cdot \sum_k \|w\|^k = \frac{\|w\|}{1-\|w\|}$ for $\|w\| < 1$, which evidently converges

[4] This means that $x < y$ implies $\exp(x) < \exp(y)$. In calculus this is also called a *strictly monotonous* map.

to 0 as $\|w\| \to 0$. Now, clearly $\exp(\overline{z}) = \overline{\exp(z)}$. Therefore, for $\theta \in \mathbb{R}$, we have $\frac{1}{\exp(i\cdot\theta)} = \exp(-i \cdot \theta) = \overline{\exp(i \cdot \theta)}$, which means that we have a group homomorphism $\exp : i \cdot \mathbb{R} \to \boldsymbol{U}$. Setting the Euler equation $\exp(i \cdot \theta) = \cos(\theta) + i \cdot \sin(\theta)$ for the real an complex parts of $\exp(i \cdot \theta)$, we have $\cos(\theta)^2 + \sin(\theta)^2 = 1$, and the alternative definitions of $\cos(\theta)$ and $\sin(\theta)$ in terms of the exponential function follow immediately. The series for $\cos(\theta)$ and $\sin(\theta)$ are also visible from the real and imaginary contributions in the series expansion of the exponential function.

We now want to calculate the kernel of exp. To this end, observe that for $k \geq 2$ and $0 < \theta \leq 3$, we have $\frac{\theta^k}{k!} > \frac{\theta^{k+1}}{(k+1)!}$. Therefore, under these conditions, the series for the cosine converges by the Leibniz criterion proposition 250. Coming back to that proposition's proof, we recognize that $1 - \frac{\theta^2}{2} < \cos(\theta) < 1 - \frac{\theta^2}{2} + \frac{\theta^4}{24}$ for $0 < \theta < 3$. But then, $u = \sqrt{2}$ is the smallest zero of $1 - \frac{\theta^2}{2}$ while $v = \sqrt{6 - 2\sqrt{3}}$ is the smallest zero of $1 - \frac{\theta^2}{2} + \frac{\theta^4}{24}$ in the interval $0 < \theta < 3$. Therefore, by proposition 245, there is a zero of $\cos(\theta)$ in the interval $]u, v[$. Since $\cos(\theta)$ is continuous by the way it is derived from $\exp(i \cdot \theta)$, and since $\cos(0) = 1$, there is a smallest zero of cos, lying between u and v. Call it $\frac{\pi}{2}$. Therefore all values in $[0, 1]$ are taken for arguments θ between 0 and $\frac{\pi}{2}$ by $\cos(\theta)$. So $\exp(i \cdot \frac{\pi}{2}) = i, \exp(i \cdot \pi) = -1, \exp(i \cdot \frac{3\pi}{2}) = -i$, and $\exp(i \cdot 2\pi) = 1$. Therefore, the cosine takes all values between 1 and -1. This implies that $\exp(i \cdot \theta)$ is onto $\boldsymbol{U}$. The goniometric addition theorem from proposition 210 in volume 1 is a consequence of the group homomorphism property of exp. For $0 \leq \theta < \theta + \eta < \frac{\pi}{2}$, it yields $\cos(\theta + \eta) = \cos(\theta)\cos(\eta) - \sin(\theta)\sin(\eta) < \cos(\theta)\cos(\eta) < \cos(\theta)$, so the cosine function is strictly monotonously decreasing. So for every $x \in [0, 1]$, there is exactly one $\theta \in [0, \frac{\pi}{2}]$ such that $\cos(\theta) = x$. By $\cos(\theta)^2 + \sin(\theta)^2 = 1$, the sine function is monotonously increasing from 0 to 1 as θ moves from 0 to $\frac{\pi}{2}$. Again, by the addition theorem for the cosine function, we have $\cos(\theta + \frac{\pi}{2}) = -\sin(\theta)$. This gives us the values for $\cos(\theta)$ for the arguments in $[\frac{\pi}{2}, \pi]$: The values $\cos(\theta)$ decrease monotonously from 0 to -1 as θ moves from $\frac{\pi}{2}$ to π. By the same argumentation, from π to 2π, $\cos(\theta)$ increases monotonously from -1 to 1. All this together proves that $i \cdot 2\pi\mathbb{Z}$ is the kernel of exp.

Let us finally concentrate on the real arguments in exp. Since $e = \exp(1) > 1$, there are arbitrary large real numbers $\exp(n) = e^n$ for real arguments, and by $\exp(-n) = \frac{1}{\exp(n)}$ also arbitrary small real values for real arguments. By proposition 245, every positive real value is taken by $\exp(x)$ for $x \in \mathbb{R}$. Now, every complex number $z \neq 0$ can be written as $z = \|z\|u$, $u \in \boldsymbol{U}$. Therefore there are $x, \theta \in \mathbb{R}$, such that $\exp(x) = \|z\|$ and $\exp(i \cdot \theta) = u$. This means that $z = \|z\|u = \exp(x)\exp(i \cdot \theta) = \exp(x + i \cdot \theta)$, and we have shown that $\exp : \mathbb{C} \to \mathbb{C}^*$ is surjective. For $x \in \mathbb{R}$, we have $\exp(-x) = \frac{1}{\exp(x)}$. But for positive $x \in \mathbb{R}$, $\exp(x) > 1$. So $\exp(x) > 0$ for all $x \in \mathbb{R}$. Moreover, for real numbers $x < y$, we have $\exp(x) < \exp(x)\exp(y - x) = \exp(y)$, whence $\exp|_{\mathbb{R}}$ is strictly monotonous onto the multiplicative group $\mathbb{R}_+$ of positive real numbers. The statements about the logarithm are now immediate. The statements about the coincidence $\exp(\frac{p}{q}) = e^{\frac{p}{q}}$ are left as an exercise. For the equation $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$, we refer to [14]. □

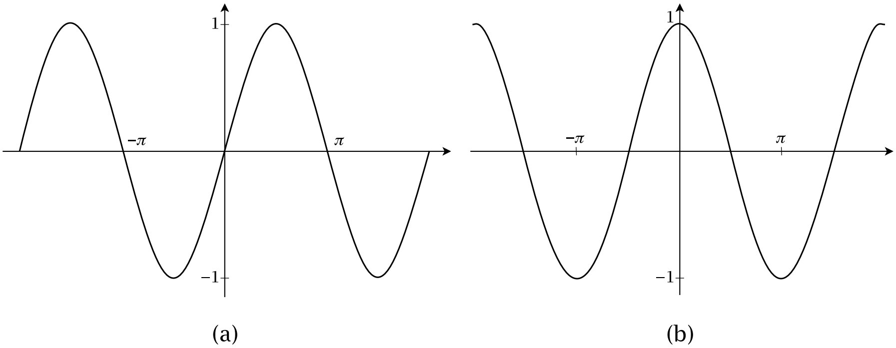

Fig. 27.10. The sine (a) and cosine (b) functions, with their domains restricted to $\mathbb{R}$.

Definition 187 *If $a \in \mathbb{R}_+$, one defines the* exponential function for basis *a by $a^x = \exp_a(x) = \exp(x \cdot \log(a))$. If moreover $a \neq 1$, one also defines the* logarithm function for basis *a by $\log_a(x) = \frac{1}{\log(a)} \log(x)$. In older literature,* $\log$ *is also denoted by* $\ln$ *(logarithmus naturalis), while one uses the notation* $\log$ *for* $\log_{10}$ *and calls that the* decadic logarithm, *but we refrain from such atavisms.*

Sorite 258 *The logarithm $\log_a$ for basis $a \in \mathbb{R}_+$ has the following properties. Let $x, y \in \mathbb{R}$.*

(i) *If $b \in \mathbb{R}_+$ is a second basis, we have $\log_b(x) = \log_b(a) \cdot \log_a(x)$,*

(ii) $\log_b(a) \cdot \log_a(b) = 1$,

(iii) $\log_a(b^x) = x \cdot \log_a(b)$,

(iv) *if $x \in \mathbb{Q}$, then the exponential function a^x and the rational powers defined earlier, denoted by the same signs, coincide,*

(v) *$a^{x+y} = a^x \cdot a^y$, and $(a^x)^y = a^{x \cdot y}$.*

(vi) *If $b \in \mathbb{R}_+$ is a second basis, we have $b^x = a^{\log_a(b) \cdot x}$.*

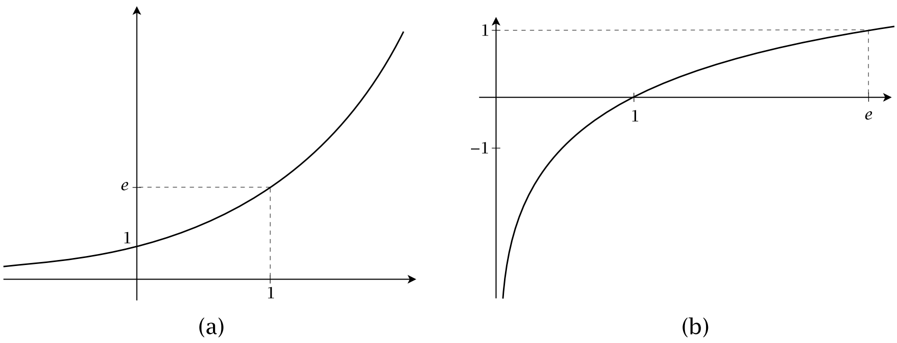

Fig. 27.11. The exponential (a) and logarithm (b) functions, with their domains restricted to $\mathbb{R}$.

Proof The proof of this sorite is left as an exercise, using the straightforward definition of the logarithm, i.e., applying the exponential function as the inverse isomorphism to log to verify the specific claims. $\square$

27.5 Proof of Euler's Formula for Polyhedra and Kuratowski's Planarity Theorem

Recall from chapter 13 of volume 1 that a skeletal graph $\Gamma : A \to {}^2V$ is a graph without multiple edges or loops. A drawing of a skeletal graph Γ as defined in definition 84, chapter 13, volume 1, is intuitively a family of (continuous) curves $c_a : [0,1] \to \mathbb{R}$ such that $c_a(]0,1[)$ is disjoint of the image of all other curves. Recall also that a drawing may also be defined on the unit sphere $S^2 \subset \mathbb{R}^3$, instead of $\mathbb{R}^2$. The *Northpole* is the top point with coordinates $(0,0,1)$. By the *stereographic projection* $\tau : S^2 - Northpole \xrightarrow{\sim} \mathbb{R}^2$, which is a homeomorphism of topological spaces, every drawing on S^2 induces one in $\mathbb{R}^2$, and conversely. Here is the definition of τ (see figure 27.12). We write $x = (h, v) \in \mathbb{R}^2 \times \mathbb{R}$ for a point in S^2.

$$\tau(h, v) = \frac{1}{1-v}h.$$

Exercise 143 Show that inverse map is $\tau^{-1}(z) = \left(\frac{2}{\|z\|^2+1} \cdot z, \frac{\|z\|^2-1}{\|z\|^2+1}\right)$. Use propositions 240 and 241 to show that these maps are continuous.

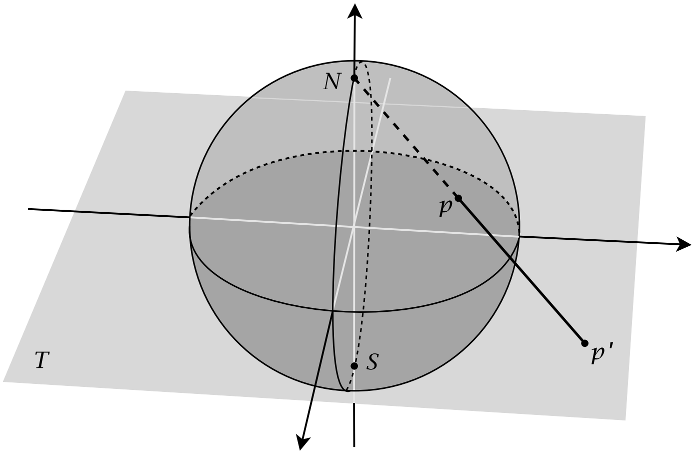

Fig. 27.12. A stereographic projection of a sphere onto the plane T through the equator. The point p is mapped to p' on T, where p, p' and the northpole N are collinear.

We called a polyhedron a drawing of a connected skeletal graph Γ on S^2, but we also use this terminology for a drawing D of Γ in $\mathbb{R}^2$, too. The elements of the finite set C of connected components of the drawn graph $D(\Gamma)$ are called the *faces of* $D(\Gamma)$. In the drawing on $\mathbb{R}^2$ there is one face which is not bounded, this one corresponds to the face on S^2 which includes the *Northpole*. It is called the *exterior face*, the others are called the *interior faces* of the drawing. We now want to prove Euler's formula for polyhedra from proposition 108, volume 1. Recall from that proposition that $\varepsilon = card(V)$, $\phi = card(A)$, $\sigma = card(C)$.

Proof The proof is by induction on the number $\xi = \varepsilon + \phi$. For $\xi = 1$, there is a single point and no edge, whence Euler's formula for polyhedra $\varepsilon - \phi + \sigma = 1 + 0 + 1 = 2$. Suppose that the drawing D has a "bridge", i.e., an edge line c_a such that the drawing minus this line is no more connected (see figure 27.13 (a)). Then the drawing of the remainder of the graph after omitting a decomposes into a disjoint union of two connected subdrawings D', D''. These subdrawings obviously each have a ξ which is smaller than that of the drawing D. Therefore, Euler's formula for polyhedra holds for both D' and D'': $\varepsilon' - \phi' + \sigma' = 2$ and $\varepsilon'' - \phi'' + \sigma'' = 2$. Now let us express ε, ϕ, and σ of D in terms of the values for D' and D'':

$$\varepsilon = \varepsilon' + \varepsilon''$$
$$\phi = \phi' + \phi'' + 1 \quad \text{(the bridge)}$$
$$\sigma = \sigma' + \sigma'' - 1 \quad \text{(both share the exterior face)}$$

So we have

$$\begin{aligned}\varepsilon - \phi + \sigma &= (\varepsilon' + \varepsilon'') - (\phi' + \phi'' + 1) + (\sigma' + \sigma'' - 1)\\ &= \varepsilon' - \phi + \sigma' + \varepsilon'' - \phi'' + \sigma'' - 1 - 1\\ &= 2 + 2 - 2\\ &= 2\end{aligned}$$

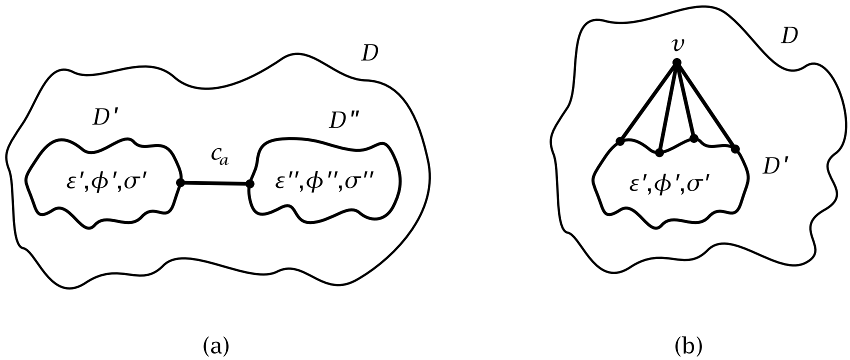

Fig. 27.13. Reducing the drawing D of a graph to drawings of smaller graphs: (a) by removing a connecting edge, (b) by removing a vertex and the edges connected to it.

If there is no bridge, take a vertex v which is on the boundary of the exterior face. The k lines terminating at v define a total of k faces containing v in their boundaries (see figure 27.13 (b)). Omitting the point v and all lines terminating in v in the drawing D defines the drawing D' of a connected graph with $k - 1$ less faces, since the interior faces around v are now united to the exterior face of D. Again, Euler's formula for polyhedra holds for D', and we have

$$\varepsilon = \varepsilon' + 1 \quad \text{(the vertex } v\text{)}$$
$$\phi = \phi' + k \quad (k \text{ edges connecting to } v)$$
$$\sigma = \sigma' + (k - 1) \quad (k - 1 \text{ new faces})$$

This yields

$$\varepsilon - \phi + \sigma = (\varepsilon' + 1) - (\phi' + k) + (\sigma' + k - 1) = 2,$$

and the proof is complete. □

Corollary 259 *The graphs K_5 and $K_{3,3}$ are not planar.*

Proof In fact, suppose that we have a drawing D of K_5. Then every three vertexes define a triangular face, so $\sigma = \binom{5}{3} = 10$, but then $\varepsilon - \phi + \sigma = 5 - 10 + 10 = 5 \neq 2$. For $K_{3,3}$, the faces are defined by rectangular cycles through 4 vertexes each. There are 9 such cycles (choose 2 upper and 2 lower points and connect them to a cycle), so we have $\varepsilon - \phi + \sigma = 6 - 9 + 9 = 6 \neq 2$ □

Now, Kuratowski's theorem attributes a central role to the two graphs K_5 and $K_{3,3}$ in that planarity of any skeletal graph Γ is based on the non-inclusion of essentially one of these non-planar graphs. "Essentially" means one of these alternatives: (1) There is a subgraph $\Gamma' \subset \Gamma$ which has a contraction isomorphic to K_5 or $K_{3,3}$ (see definition 85, chapter 13, vol 1). (2) There is a subgraph $\Gamma' \subset \Gamma$ which results from K_5 or $K_{3,3}$ by a succession of subdivisions of their edges. A subdivisions of an edge $x \xrightarrow{a} y$ is the addition of one more vertex v to V and the replacement of a by two edges $x \xrightarrow{a_x} v$ and $v \xrightarrow{a_y} y$.

We have these auxiliary facts:

Lemma 260 *Suppose that we can prove the special case that Γ is planar iff it contains no subgraph which is a subdivision of a graph isomorphic to K_5 or $K_{3,3}$. Then the theorem follows.*

Proof If Γ has no subgraph which can be contracted to a graph isomorphic to K_5 or $K_{3,3}$, then in particular, it has no subgraph, which is a subdivision of a graph isomorphic to K_5 or $K_{3,3}$, since subdivisions can be contracted to the original graphs. By the assumption made in lemma 260 it then can be concluded that Γ is planar. Conversely, if Γ is planar and there is a contraction to a graph isomorphic to K_5 or $K_{3,3}$, then there is a sequence of elementary contractions, which define this contraction. The idea is this: If it can be shown that an elementary contraction preserves the planarity of a graph, then K_5 or $K_{3,3}$ must be planar, which a contradiction. So, if a drawing of a planar graph Γ is given, an elementary contraction of the line $x \xrightarrow{a} y$ can be performed by isolating a small tubular neighborhood around the drawing of a and then piping the lines ending at x within that tubular neighborhood to y (see figure 27.14). Obviously, this construction conserves planarity. Thus the lemma is proved. □

Kuratowski's theorem

So one is left with the proof of the subdivision version of Kuratowski's theorem. Now, we already know that a graph containing a subdivision of

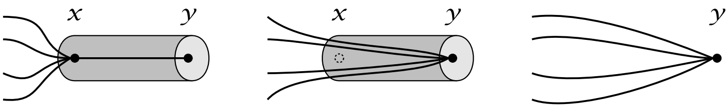

Fig. 27.14. The graphical process of an elementary contraction conserves planarity.

a copy of K_5 or of $K_{3,3}$ can not be planar, since the contractions yielding K_5 or $K_{3,3}$ would yield drawings of graphs containing drawings of K_5 or of $K_{3,3}$, which is impossible by corollary 259. So we are left with the proof of the other implication, i.e., that a non-planar graph must necessarily contain a subdivision of drawings of K_5 or of $K_{3,3}$.

Suppose there is a Γ which, being non-planar, contains no subgraph which is a subdivision of a copy of K_5 or of $K_{3,3}$. Take one with a minimal number of edges. It cannot have a bridge line, since then it is easily seen that one of the subgraphs connected by this bridge would be non-planar and therefore would contain a subdivision of one of the two critical graphs. Moreover, it cannot contain points x with $deg(x) = 1, 2$, since the non-planarity would be conserved omitting these points. So all points have $deg(x) \geq 3$. Then the omission of an arbitrary line $x \overset{l}{\text{---}} y$ in Γ yields a smaller graph Φ which does not contain a subdivision of K_5 or of $K_{3,3}$ and therefore is planar.

The proof idea is to show that, under these assumptions, one can find a subgraph of Φ which is isomorphic to K_5 or $K_{3,3}$, and this would contradict the assumption that Γ does not contain any of these subgraphs. To do so, one first shows that there is a cycle Z in Φ containing the points x, y defined above. One then makes a drawing of Φ such that there is a maximum of faces interior to the drawing of Z. One considers the components of the subgraph of Φ induced on the vertexes outside the drawing of Z and then defines *outer pieces* as those subgraphs of Φ which are either induced on outer components, plus the points on Z which they are connected to, or else which are outer edges of the drawing of Z connecting two points of Z. Inner components and inner pieces are defined in an analogous way. For a pair of points u, v on Z, one looks for inner or outer pieces such that they contain points $\neq u, v$ on both walks on Z (in clockwise orientation, say) between u and v. These pieces are called *$(u$–$v)$-separating*.

One can find an inner piece H and four points u_0, u_1, v_0, v_1 such that

$$Z = u_0 \text{——} \ldots u_1 \text{——} \ldots v_0 \text{——} \ldots v_1 \text{——} \ldots u_0,$$

and such that H is (u_0–v_0)- and (u_1–v_1)-separating. Thus H meets the clockwise walks $u_0 \text{——} \ldots u_1$, $u_1 \text{——} \ldots v_0$, $v_0 \text{——} \ldots v_1$, and $v_1 \text{——} \ldots u_0$ in four points q, r, s, t, all different from u_0, u_1, v_0, v_1. The proof now closes with an analysis of four cases of possible positions of the points q, r, s, t on the cycle Z, and where each case yields a subgraph isomorphic to K_5 or $K_{3,3}$. This is a contradiction to the assumption that the original graph Γ (of which Φ is a subgraph) does not contain a subgraph isomorphic to K_5 or $K_{3,3}$. The details of the proof are described in [12]. It goes back to Gabriel Andrew Dirac and Seymour Schuster, A theorem of Kuratowski. Nderl. Akad. Wetensch. Proc. Ser. A 57, 1954.