Chapter 2
Topology

The first four sections of this chapter contain a brief summary of results of analysis most theoretical physicists are more or less familiar with.

2.1 Basic Definitions

A **topological space** is a double $(X, \mathcal{T})$ of a set X and a family $\mathcal{T}$ of subsets of X specified as the **open sets** of X with the following properties:

1. $\emptyset \in \mathcal{T},\ X \in \mathcal{T} \quad (\emptyset$ is the empty set),
2. $(\mathcal{U} \subset \mathcal{T}) \Rightarrow \left(\bigcup_{U \in \mathcal{U}} U \in \mathcal{T} \right)$,
3. $(U_n \in \mathcal{T}$ for $1 \leq n \leq N \in \mathbb{N}) \Rightarrow \left(\bigcap_{n=1}^{N} U_n \in \mathcal{T} \right)$,

that is, $\mathcal{T}$ is closed under unions and under finite intersections. If there is no doubt about the family $\mathcal{T}$, the topological space is simply denoted by X instead of $(X, \mathcal{T})$.

Two **topologies** $\mathcal{T}_1$ and $\mathcal{T}_2$ on X may be compared, if one is a subset of the other; if $\mathcal{T}_1 \subset \mathcal{T}_2$, then $\mathcal{T}_1$ is **coarser** than $\mathcal{T}_2$ and $\mathcal{T}_2$ is **finer** that $\mathcal{T}_1$. The coarsest topology is the **trivial topology** $\mathcal{T}_0 = \{\emptyset, X\}$, the finest topology is the **discrete topology** consisting of all subsets of X.

A **neighborhood** of a point $x \in X$ (of a set $A \subset X$) is an open[1] set $U \in \mathcal{T}$ containing x as a point (A as a subset). The complements $C = X \setminus U$ of open sets $U \in \mathcal{T}$ are the **closed sets** of the topological space X. If $A \in X$ is any set, then the **closure** $\overline{A}$ of A is the smallest closed set containing A, and the **interior** $\mathring{A}$ of A is the largest open set contained in A; $\overline{A}$ and $\mathring{A}$ always exist by Zorn's lemma. $\mathring{A}$ is

[1] In this text neighborhoods are assumed open; more generally a neighborhood is any set containing an open neighborhood.

H. Eschrig, *Topology and Geometry for Physics*, Lecture Notes in Physics, 822,
DOI: 10.1007/978-3-642-14700-5_2, © Springer-Verlag Berlin Heidelberg 2011

the set of **inner points** of A. $\overline{A}$ is the set of **points of closure** of A; points every neighborhood of which contains at least one point of A. (The complement of $\overline{A}$ is the largest open set not intersecting A.) The **boundary** ∂A of A is the set $\overline{A} \setminus \mathring{A}$. A is **dense** in X, if $\overline{A} = X$. A is **nowhere dense** in X, if the interior of $\overline{A}$ is empty: $(\overline{A})^\circ = \emptyset$. X is **separable** if $X = \overline{A}$ for some *countable* set A.

(One might wonder about the asymmetry of axioms 2 and 3. However, if closure under all intersections would be demanded, no useful theory would result. For instance, a point of the real line $\mathbb{R}$ can be obtained as the intersection of an infinite series of open intervals. Hence, with the considered modification of axiom 3, points and all subsets of $\mathbb{R}$ would be open and closed and the topology would be discrete as soon as all open intervals are open sets.)

The **relative topology** $\mathcal{T}_A$ on a subset A of a topological space $(X, \mathcal{T})$ is $\mathcal{T}_A = \{A \cap T | T \in \mathcal{T}\}$, that is, its open sets are the intersections of A with open sets of X. Consider the closed interval $[0, 1]$ on the real line $\mathbb{R}$ with the usual topology of unions of open intervals on $\mathbb{R}$. The half-open interval $]x, 1], 0 < x < 1$, of $\mathbb{R}$ is an open set in the relative topology on $[0, 1] \subset \mathbb{R}$!

Most of the interesting topological spaces are **Hausdorff**: any two distinct points have disjoint neighborhoods. (A non-empty space of at least two points and with the trivial topology is not Hausdorff.) In a Hausdorff space single point sets $\{x\}$ are closed. (Exercise, take neighborhoods of all points distinct from x.)

Sequences are not an essential subject in this book. Just to be mentioned, a sequence of points in a topological space X converges to a point x, if every neighborhood of x contains all but finitely many points of the sequence. A partially ordered set I is directed, if every pair a, b of elements of I has an upper bound $c \in I, c \geq a, c \geq b$. A set of points of X is a net, if it is indexed by a directed index set I. A net converges to a point x, if for every neighborhood U of x there is an index b so that $x_a \in U$ for all $a \geq b$. In Hausdorff spaces points of convergence are unique if they exist.

The central issue of topology is continuity. A function (mapping) f from a topological space X into a topological space Y (maybe the same space X) is **continuous** at $x \in X$, if given any (in particular small) neighborhood V of $f(x) \subset Y$ there is a neighborhood U of x such that $f(U) \subset V$ (compare Fig. 1.1 of Chap. 1). The function f is continuous if it is continuous at every point of its domain. In this case, the inverse image $f^{-1}(V)$ of any open set V of the target space Y of f is an open set of X. (It may be empty.) The coarser the topology of Y or the finer the topology of X the more functions from X into Y are continuous. Observe that, *if X is provided with the discrete topology, then every function $f : X \to Y$ is continuous, no matter what the topology of Y is*. If $f : X \to Y$ and $g : Y \to Z$ are continuous functions, then their composition $g \circ f : X \to Z$ is obviously again a continuous function.

Consider functions $f(x) = y : [0, 1] \to \mathbb{R}$. What means continuity at $x = 1$ if the relative topology of $[0, 1] \subset \mathbb{R}$ is taken?

f is continuous iff it maps convergent nets to convergent nets; in metric spaces sequences suffice instead of nets.

A **homeomorphism** is a bicontinuous bijection f (f and f^{-1} are continuous functions *onto*); it maps open sets to open sets and closed sets to closed sets. A homeomorphism from a topological space X to a topological space Y provides a one–one mapping of points and a one–one mapping of open sets, hence it provides an equivalence relation between topological spaces; X and Y are called **homeomorphic**, $X \sim Y$, if a homeomorphism from X to Y exists. There exists always the identical homeomorphism Id_X from X to X, and a composition of homeomorphisms is a homeomorphism. The topological spaces form a category the morphisms of which are the continuous functions and the isomorphisms are the homeomorphisms (see Compendium C.1 at the end of the book).

A **topological invariant** is a property of topological spaces which is preserved under homeomorphisms.

2.2 Base of Topology, Metric, Norm

If topological problems are to be solved, it is in most cases of great help that not the whole family $\mathcal{T}$ of a topological space $(X, \mathcal{T})$ need be considered.

A subfamily $\mathcal{B}$ of $\mathcal{T}$ is called a **base of the topology** $\mathcal{T}$ if every $U \in \mathcal{T}$ can be formed as $U = \cup_\beta B_\beta,\ B_\beta \in \mathcal{B}$. A family $\mathcal{B}(x)$ is called a **neighborhood base** at x if each $B \in \mathcal{B}(x)$ is a neighborhood of x and given any neighborhood U of x there is a B with $U \supset B \in \mathcal{B}(x)$. A topological space is called **first countable** if each of its points has a countable neighborhood base, it is called **second countable** if it has a countable base.

The **product topology** on the product $X \times Y$ of topological spaces X and Y is defined by the base consisting of sets

$$\{(x, y) | x \in B_X,\ y \in B_Y\}, \quad B_X \in \mathcal{B}_X, \quad B_Y \in \mathcal{B}_Y, \tag{2.1}$$

where $\mathcal{B}_X$ and $\mathcal{B}_Y$ are bases of topology of X and Y, respectively. It is the coarsest topology for which the canonical projection mappings $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are continuous (exercise). The $\mathbb{R}^n$ with its usual topology is the topological product $\mathbb{R} \times \cdots \times \mathbb{R}, n$ times.

A very frequent special case of topological space is a metric space. A set X is a **metric space** if a non-negative real valued function, the **distance function** $d : X \times X \to \mathbb{R}_+$ is given with the following properties:

1. $d(x, y) = 0, \quad \text{iff}\, x = y,$
2. $d(x, y) = d(y, x),$
3. $d(x, z) \leq d(x, y) + d(y, z) \quad$ (triangle inequality).

An **open ball** of radius r with its center at point $x \in X$ is defined as $B_r(x) = \{x' | d(x, x') < r\}$. The class of all open balls forms a base of a topology of X, the **metric topology**. It is Hausdorff and first countable; a neighborhood base of point x is for instance the sequence $B_{1/n}(x),\ n = 1, 2, \ldots$

The metric topology is uniquely defined by the metric as any topology is uniquely defined by a base. There are, however, in general many different metrics defining the same topology. For instance, in $\mathbb{R}^2 \ni x = (x^1, x^2)$ the metrics

$$d_1(\boldsymbol{x},\boldsymbol{y}) = ((x^1 - y^1)^2 + (x^2 - y^2)^2)^{1/2} \quad \text{Euclidean metric,}$$
$$d_2(\boldsymbol{x},\boldsymbol{y}) = \max\{|x^1 - y^1|, |x^2 - y^2|\},$$
$$d_3(\boldsymbol{x},\boldsymbol{y}) = |x^1 - y^1| + |x^2 - y^2| \quad \text{Manhattan metric}$$

define the same topology (exercise).

A sequence $\{x_n\}$ in a metric space is **Cauchy** if

$$\lim_{m,n\to\infty} d(x_m, x_n) = 0. \tag{2.2}$$

A metric space X is **complete** if every Cauchy sequence converges in X (in the metric topology). The rational line $\mathbb{Q}$ is not complete, the real line $\mathbb{R}$ is, it is an isometric completion of $\mathbb{Q}$. An **isometric completion** $\tilde{X}$ of a metric space X always exists in the sense that $\tilde{X} \supset X$ is complete, $\tilde{X} = \overline{X}$ (closure of X in $\tilde{X}$), and the distance function $d(x, x')$ is extended to $\tilde{X}$ by continuity. $\tilde{X}$ is unique up to isometries (distance preserving transformations) which leave the points of X on place. A complete metric space is a **Baire space**, that is, it is not a countable union of nowhere dense subsets. The relevance of this statement lies in the fact that if a complete metric space is a countable union $X = \cup_n U_n$, then some of the U_n must have a non-empty interior [1, Section III.5].

A metric space X is complete, iff every sequence $C_1 \supset C_2 \supset \dots$ of closed balls with radii $r_1, r_2, \dots \to 0$ has a non-empty intersection.

Proof Necessity: Let X be complete. The centers x_n of the balls C_n obviously form a Cauchy sequence which converges to some point x, and $x \in \cap_n C_n$. Sufficiency: Let x_n be Cauchy. Pick n_1 so that $d(x_n, x_{n_1}) < 1/2$ for all $n \geq n_1$ and take x_{n_1} as the center of a ball C_1 of radius $r_1 = 1$. Pick $n_2 \geq n_1$ so that $d(x_n, x_{n_2}) < 1/2^2$ for all $n \geq n_2$ and take x_{n_2} as the center of a ball C_2 of radius $r_2 = 1/2$... The sequence $C_1 \supset C_2 \supset \dots$ has a non-empty intersection containing some point x. It is easily seen that $x = \lim x_n$. □

Let X be a metric space and let $F : X \to X : x \mapsto Fx$ be a **strict contraction**, that is a mapping of X into itself with the property

$$d(Fx, Fx') \leq \lambda d(x, x'), \quad \lambda < 1. \tag{2.3}$$

(A contraction is a mapping which obeys the weaker condition $d(Fx, Fx') \leq d(x, x')$; every contraction is obviously continuous since the preimage of any open ball $B_r(Fx)$ contains the open ball $B_r(x)$. Exercise.) A vast variety of physical problems implies **fixed point equations**, equations of the type $x = Fx$. Banach's contraction mapping principle says that *a strict contraction F on a complete metric space X has a unique fixed point.*

Proof Uniqueness: Let $x = Fx$ and $y = Fy$, then $d(x, y) = d(Fx, Fy) \leq \lambda d(x, y)$, $\lambda < 1$. Hence, $d(x, y) = 0$ that is $x = y$. Existence: Pick x_0 and let $x_n = F^n x_0$. Then, $d(x_{n+1}, x_n) = d(Fx_n, Fx_{n-1}) \leq \lambda d(x_n, x_{n-1}) \leq \dots \leq \lambda^n d(x_1, x_0)$. Thus, if $n > m$, by the triangle inequality and by the sum of a geometrical series,

$d(x_n, x_m) \leq \sum_{l=m+1}^{n} d(x_l, x_l - 1) \leq \lambda^m (1-\lambda)^{-1} d(x_1, x_0) \to 0$ for $m, n \to \infty$ implying that $\{x_n\} = \{Fx_{n-1}\}$ is Cauchy and converges towards an $x \in X$. By continuity of $F, x = Fx$. □

Equation systems, systems of differential equations, integral equations or more complex equations may be cast into the form of a fixed point equation. A simple case is the equation $x = f(x)$ for a function $f : [a, b] \to [a, b]$, $[a, b] \subset \mathbb{R}$, obeying the Lipschitz condition

$$|f(x) - f(x')| \leq \lambda |x - x'|, \quad \lambda < 1, \quad x, x' \in [a, b].$$

If for instance $|f'(x)| \leq \lambda < 1$ for $x \in [a, b]$, the Lipschitz condition is fulfilled. From Fig. 2.1 it is clearly seen how the solution process $x_n = f(x_{n-1})$ converges. The convergence is fast if $|f'(x)| \ll 1$. Consider this process for $|f'(x)| > 1$. Next consider $a = -\infty$; why is a simple contraction not sufficient and a strict contraction needed to guarantee the existence of a solution?

There are always many ways to cast a problem into a fixed point equation. If $x = Fx$ has a solution x_0, it is easily seen that $x = \tilde{F}x$ with $\tilde{F}x = x + p(Fx - x)$ has the same solution x_0. If F is not a strict contraction, $\tilde{F}$ with a properly chosen p sometimes is, although possibly with a very slow convergence of the solution process. Sophisticated constructions have been developed to enforce convergence of the solution process of a fixed point equation.

Another frequent special case of topological space is a **topological vector space** X over a field K. (In most cases $K = \mathbb{R}$ or $K = \mathbb{C}$.) It is also a vector space (see Compendium) and its topology is such that the mappings

$$K \times X \to X : (\lambda, x) \mapsto \lambda x,$$
$$X \times X \to X : (x, x') \mapsto x + x'$$

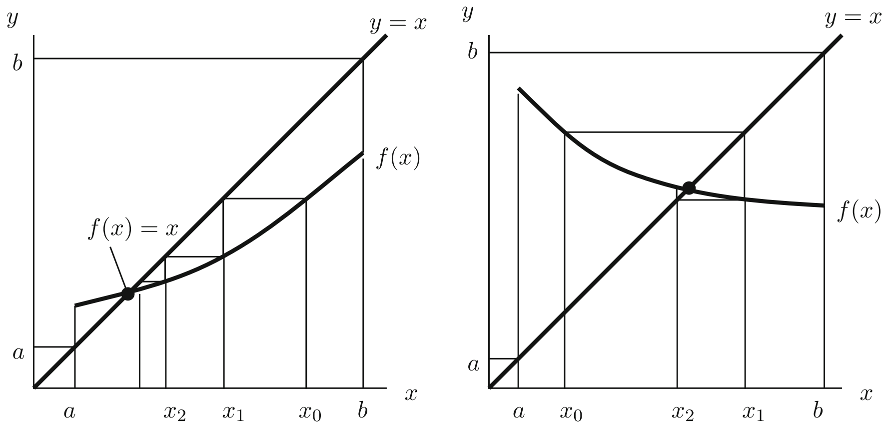

Fig. 2.1 Illustration of the fixed point equation $x = f(x)$ for $f'(x) > 0$ (*left*) and $f'(x) < 0$ (*right*)

are continuous, where K is taken with its usual metric topology and $K \times X$ and $X \times X$ are taken with the product topology. If $\mathcal{B}(0)$ is a neighborhood base at the origin of the vector space X, then the set $\mathcal{B}(x)$ of all open sets $B_\beta(x) = x + B_\beta(0) = \{x + x' | x' \in B_\beta(0)\}$ with $B_\beta(0) \in \mathcal{B}(0)$ is a neighborhood base at x. For any open (closed) set A, $x + A$ is open (closed). For two sets $A \subset X, B \subset X$ the vector sum is defined as $A + B = \{x + x' | x \in A, x' \in B\}$.

Linear independence of a set $E \subset X$ means that if $\sum_{n=1}^{N} \lambda^n x_n = 0$ (upper index at λ^n, not power of λ) holds for any finite set of N *distinct* vectors $x_n \in E$, then $\lambda^n = 0$ for all $n = 1, \ldots, N$. Linear independence (as well as its opposite, linear dependence) is a property of the algebraic structure of the vector space, not of its topology. A **base** E in a topological vector space is a linearly independent subset the **span** of which (the set of all linear combinations over K of finitely many vectors out of E) is dense in $X : \overline{\mathrm{span}_K E} = X$. It is a base of vector space, not a base of topology. It may, however, depend on the topology of X. The maximal number of linearly independent vectors in E is the **dimension** of the topological vector space X; it is a finite integer n or infinity, countable or not. If the dimension of a topological vector space X is $n < \infty$, then X is homeomorphic to K^n. If it is infinite, the dimension is to be distinguished from the algebraic dimension of the vector space (see Compendium). It can be shown that a topological vector space X is **separable** if it admits a countable base. Any vector x of $\mathrm{span}_K E$ has a unique representation $x = \sum_{n=1}^{N} \lambda^n x_n$, $x_n \in E$ with some finite N. Hence, if X is Hausdorff, then every vector $x \in X$ has a unique representation by a converging series $x = \sum_{n=1}^{\infty} \lambda^n x_n$, $x_n \in E$ (exercise).

Two subspaces (see Compendium) M and N of a vector space X are called **algebraically complementary**, if $M \cap N = \{0\}$ and $M + N = X$. X is then said to be the **direct sum** $M \oplus N$ of the vector spaces M and N. Consider all possible sets $x + M$, $x \in X$. They either are disjoint or identical (exercise). Let $\tilde{x}$ be the equivalence class of the set $x + M$. By an obvious canonical transfer of the linear structure of X into the set of classes $\tilde{x}$ these classes form a vector space; it is called the **quotient space** X/M of X by M (Fig. 2.2). Let the topology of X be such that the one point set $\{0\}$ is closed. Then, for any $x \in X$, $M_x = \{\lambda x | \lambda \in K\}$ is a closed subspace of X (exercise).

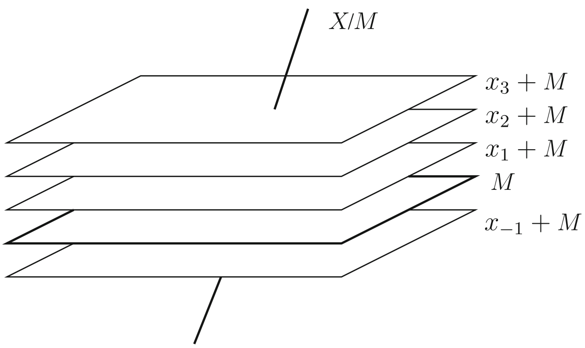

Fig. 2.2 A subspace M of a vector space X and cosets $x_i + M$ with x_i linearly independent of M. Note that an angle between X/M and M has no meaning so far

It is just by custom that the cosets $x_i + M$ were drawn as parallel planes in Fig. 2.2, and that X/M was drawn as a straight line. Angles, curvature and all that is not defined as long as X is considered as a topological vector space only. Any continuous deformation of Fig. 2.2 is admitted. Even if a metric is defined on a one-dimensional vector space, say, it would not make a difference if it would be drawn as a straight line or a spiral provided it is consistently declared how to relate the point λx to the point x. These remarks are essential in later considerations.

A topological vector space X is said to be **metrizable** if its topology can be deduced from a metric that is translational invariant: $d(x, x') = d(x + a, x' + a)$ for all $a \in X$. Many topological vector spaces, in particular all metrizable vector spaces, are **locally convex**: they admit a base of topology made of convex sets. (A set of a vector space is convex if it contains the 'chord' between any two of its points, that is, if x and x' are two points of the set then all points $\lambda x + (1 - \lambda)x'$, $0 < \lambda < 1$ belong to the set.)

In most cases a metrizable topological vector space is metrized either by a family of seminorms or by a norm. A **norm** is a real function $x \mapsto ||x||$ with the properties

1. $||x + x'|| \leq ||x|| + ||x'||$,
2. $||\lambda x|| = |\lambda| \, ||x||$,
3. $||x|| = 0, \ \text{iff}\, x = 0$.

From the first two properties the non-negativity of a norm follows; if the last property is abandoned one speaks of a **seminorm**.The metric of a norm is given by $d(x, x') = ||x - x'||$. A complete metrizable vector space is a **Fréchet space**, a complete normed vector space is a **Banach space**. Fréchet spaces whose metric does not come from a single norm are used in the theory of generalized functions (distributions).

A **linear function (operator)** $L : X \to Y$ from a vector space X into a vector space Y over the same field K is a function with the property

$$L(\lambda x + \lambda' x') = \lambda L(x) + \lambda' L(x'), \quad \lambda, \lambda' \in K. \tag{2.4}$$

A function from a vector space X into its field of scalars K is called a **functional**, if it is linear it is called a **linear functional**. A linear function from a topological vector space into a topological vector space is continuous, iff it is continuous at the origin $x = 0$ (exercise). A linear function from a normed vector space X into a normed vector space Y (for instance the one-dimensional vector space K) is **bounded** if

$$||L|| = \sup_{0 \neq x \in X} \frac{||L(x)||_Y}{||x||_X} < \infty. \tag{2.5}$$

The operator notation Lx is often used instead of $L(x)$. A linear function from a normed vector space into a normed vector space is bounded, iff it is continuous (exercise). With the norm (2.5) (prove that it is indeed a norm), the set $\mathcal{L}(X, Y)$ of all bounded linear operators with linear operations among them defined in the natural way is again a normed vector space; *it is Banach if Y is Banach.*

Proof Let $\{L_n\}$ be Cauchy. Since $|||L_m|| - ||L_n||| \leq ||L_m - L_n|| \to 0$, $\{||L_n||\}$ is a Cauchy sequence of real numbers converging to some real number C. For each $x \in X$, $\{L_n x\}$ is a Cauchy sequence in Y. Since Y is complete, $L_n x$ converges to some point $y \in Y$. Define L by $Lx = y$. Then, $||Lx|| = \lim_{n\to\infty} ||L_n x|| \leq \lim_{n\to\infty} ||L_n||\,||x|| = C||x||$, where (2.5) was used. Hence, L is a bounded operator. Moreover, $||(L - L_n)x|| = \lim_{m\to\infty} ||(L_m - L_n)x|| \leq \lim_{m\to\infty} ||(L_m - L_n)||||x||$ and therefore $\lim_{n\to\infty} ||L - L_n|| = \lim_{n\to\infty} \sup_{x\neq 0} ||(L - L_n)x||/||x|| \leq \lim_{m,n\to\infty} ||L_m - L_n|| = 0$. Hence, L_n converges to L in the operator norm. □

The **topological dual** X^* of a topological vector space X is the set of all continuous **linear functionals**

$$f : X \to K : x \mapsto \langle f, x\rangle \in K, \quad \langle f, \lambda x + \lambda' x'\rangle = \lambda\langle f, x\rangle + \lambda'\langle f, x'\rangle, \tag{2.6}$$

from X into K provided with the natural linear structure $\langle \lambda f + \lambda' f', x\rangle = \lambda\langle f, x\rangle + \lambda'\langle f', x\rangle$. It is again a normed vector space with the norm $||f||$ given by (2.5) with f instead of L, $||f|| = \sup_{0\neq x\in X} |\langle f, x\rangle|/||x||_X$. As there are the less continuous functions the coarser the topology of the domain space is, the question arises, what is the coarsest topology of X for which all bounded linear functionals are continuous. This topology of X is called the **weak topology**. A neighborhood base of the origin for this weak topology is given by all intersections of finitely many open sets $\{x|\,|\langle f, x\rangle| < 1/k\}$, $k = 1, 2\ldots$ for all $f \in E^*$, a base of the vector space X^*. For instance, if $X = \mathbb{R}^n$, these open sets comprise all infinite 'hyperplates' of thickness $2/k$ sandwiching the origin and normal in turn to one of the n base vectors $\boldsymbol{f}^i$ of $X^* = \mathbb{R}^n$ (Fig. 2.3). Taken for every k, the intersections of n such 'hyperplates' containing $\{0\} \in X$ form a neighborhood base of the origin of $\mathbb{R} \times \cdots \times \mathbb{R}, n$ factors, in the product topology which in this case is equivalent to the standard norm topology of $\mathbb{R}^n$. Hence, the $\mathbb{R}^n$ with both the weak and the norm topologies are homeomorphic to each other and can be identified with each other. This does not hold true for an infinite dimensional space X.

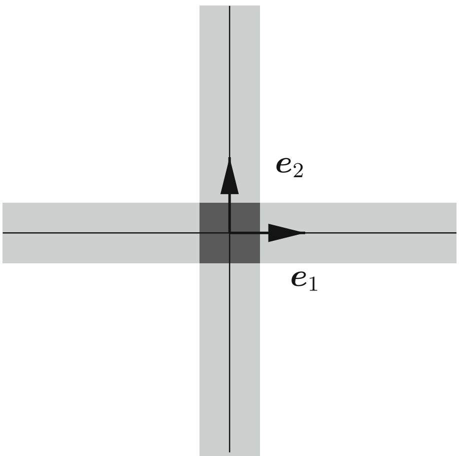

Fig. 2.3 Open sets of a neighborhood base of the origin of the $\mathbb{R}^2$ in the weak topology and their intersection

The topological dual of a normed vector space X is $X^* = \mathcal{L}(X, K)$ (with the norm $||f||$ as above); if K is complete (as $\mathbb{R}$ or $\mathbb{C}$) then X^* is a Banach space (no matter whether X is complete or not). The second dual of X is the dual $X^{**} = (X^*)^*$ of X^*. Let $J : X \to X^{**} : x \mapsto \tilde{x}$ where $\langle \tilde{x}, f \rangle = \langle f, x \rangle$ for all $f \in X^*$.

If X is a Banach space then the above mapping J is an isometric isomorphism of X onto a subspace of X^{**}, hence, one may consider $X \subset X^{**}$.

The proof of this statement makes use of the famous Hahn–Banach theorem which provides the existence of ample sets of continuous linear functionals [1, Section III.2.3]. X is said to be **reflexive**, if the above mapping J is *onto* X^{**}. In this case one may consider $X = X^{**}$.

An **inner product** (or scalar product) in a complex vector space X is a **sesquilinear function** $X \times X \to \mathbb{C} : (x, y) \mapsto (x|y)$ with the properties

1. $(x|y) = \overline{(y|x)}$,
2. $(x|y_1 + y_2) = (x|y_1) + (x|y_2)$,
3. $(x|\lambda y) = \lambda(x|y)$ (convention in physics),
4. $(x|x) > 0 \quad \text{for}\, x \neq 0$.

(In mathematics literature, the convention $(\lambda x|y) = \lambda(x|y)$ is used instead of 3.) An inner product in a real vector space X is the corresponding bilinear function $X \times X \to \mathbb{R}$ with the same properties 1 through 4. ($\bar{\lambda}$ is the complex conjugate of λ, in $\mathbb{R}$ of course $\bar{\lambda} = \lambda$.) If an inner product is given,

$$||x|| = (x|x)^{1/2} \tag{2.7}$$

has all properties of a norm (exercise, use the Schwarz inequality given below). A normed vector space with a norm of an inner product is called an **inner product space** or a pre-Hibert space. A complete inner product space is called a **Hilbert space**. Some authors call it a Hilbert space only if it is infinite-dimensional; a finite-dimensional inner product space is also called a **unitary space** in the complex case and a **Euclidean space** in the real case. Two Hilbert spaces X and X' are said to be **isomorphic** or unitarily equivalent, $X \approx X'$, if there exists a **unitary operator** $U : X \to X'$, that is, a surjective linear operator for which $(Ux|Uy) = (x|y)$ holds for all $x, y \in X$ (actually it is bijective, exercise).

In an inner product space the **Schwarz inequality**

$$|(x|y)| \leq ||x|| \, ||y|| \tag{2.8}$$

holds, and in a *real* inner product space the **angle** between vectors x and y is defined as

$$\cos(\measuredangle(x, y)) = \frac{(x|y)}{||x|| \, ||y||}. \tag{2.9}$$

Proof of the Schwarz inequality Let $\hat{y} = y/||y||$ and $x_1 = (\hat{y}|x)\hat{y}$, $x_2 = x - x_1$ implying $(x_1|x_2) = 0$, $x = x_1 + x_2$. Then, $||x||^2 = (x_1 + x_2|x_1 + x_2) = ||x_1||^2 + ||x_2||^2 \geq ||x_1||^2 = |(x|y)|^2/||y||^2$. $\square$

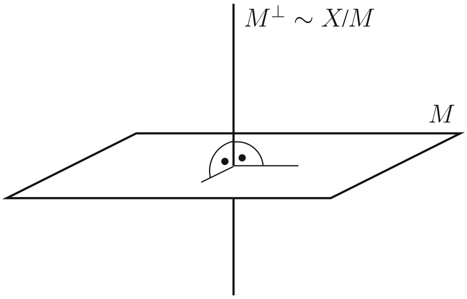

Fig. 2.4 Orthogonal complement $M^\perp$ to a closed subspace M of an inner product space

Even in a complex inner product space, orthogonality is defined: two vectors x and y are **orthogonal** to each other, if $(x|y) = 0$. An **orthonormalized base** in an inner product space is a base E (of topological vector space, see p. 16) with $||e|| = 1$ and $(e|e') = 0$, $e \neq e'$ for all $e, e' \in E$. Let $\{e_n\}_{n=1}^N \subset E$. A slight generalization of the proof of the Schwarz inequality proves **Bessel's inequality**: $||x||^2 \geq \sum_{n=1}^N |(e_n|x)|^2$. If M is a closed subspace of an inner product space X, then the set of all vectors of X which are orthogonal to all vectors of M forms the **orthogonal complement** $M^\perp$ of M in X (Fig. 2.4, compare to Fig. 2.2). Every vector $x \in X$ has a unique decomposition $x = x_1 + x_2$, $x_1 \in M$, $x_2 \in M^\perp$ (exercise), that is, $X = M + M^\perp$.

If X and X' are two Hilbert spaces over the same field K then their **direct sum** $X \oplus X'$ is defined as the set of all ordered pairs $(x, x'), x \in X, x' \in X'$ with the scalar product $((x, x')|(y, y')) = (x|y)_X + (x'|y')_{X'}$. (Hence, in the above case also $X = M \oplus M^\perp$ holds.) The direct sum of more that two, possibly infinitely many Hilbert spaces is defined accordingly. (The vectors of the latter case are the sequences $\{x^i\}$ for which the sum of squares of norms converges.)

The **tensor product** $X \otimes X'$ of Hilbert spaces X and X' is defined in the following way: Consider pairs $(x, x') \in X \times X'$ and define for each pair a bilinear function $x \otimes x'$ on the product vector space $X \times X'$ by $x \otimes x'(y, y') = (x|y)(x'|y')$. Consider the linear space of all finite linear combinations $\varphi = \sum_{n=1}^N c_n x_n \otimes x'_n$ and define an inner product $(\varphi|\psi)$ by linear extension of $(x \otimes x'|y \otimes y') = (x|y)(x'|y')$. The completion of this space is $X \otimes X'$. (Exercise: show that $(\varphi|\psi) = 0$ if $\varphi = \sum_{n=1}^N c_n x_n \otimes x'_n = 0$ and that $(\varphi|\psi)$ has the four properties of a scalar product.)

Finally, let X be a Hilbert space and let $y \in X$. Then, $f_y(x) = (y|x)$ is a continuous linear function $f_y : X \to K : x \mapsto (y|x)$, hence $f_y \in X^*$. The Riesz lemma says that there is a *conjugate* linear bijection $y \mapsto f_y$ between X and its dual X^* [1].

We close the section with a number of examples of vector spaces from physics:

$\mathbb{R}^n = \mathbb{R}^{n*}$, the set of real n-tuples $\boldsymbol{a} = \{a^1, a^2, \ldots, a^n\}$, is used as a mere topological vector space with the product topology of $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (n factors) or as a Euclidean space (real finite-dimensional Hilbert space, $(a|b) = \boldsymbol{a} \cdot \boldsymbol{b} = \sum a^i b^i$ implying the same topology) in the sequel, depending on context (cf. the discussion in connection with Figs. 2.2 and 2.4). Both concepts play a central role in the theory of real manifolds. As a mere topological vector space it is the **configuration space** of a many-particle system, as an Euclidean space the position space or the momentum space of physics. For instance in the **physics of vibrations**, $\mathbb{C}^n \approx \mathbb{R}^{2n}$ by the isomorphism $z^j = x^j + iy^j \mapsto (a^{2j-1}, a^{2j}) = (x^j, y^j)$ is used, where only the x^j describe actual amplitudes. In the sequel, vectors of the space K^n ($K = \mathbb{R}$ or $\mathbb{C}$) are denoted by bold-face letters and the inner product is denoted by a dot.

l^p as sequence spaces the points of which are complex or real number sequences $a = \{a^i\}_{i=1}^{\infty}$ are defined for $1 \le p < \infty$ with the norm ($a \in l^p$, iff $||a||_p < \infty$)

$$l^p: \ ||a||_p = \left(\sum_{i=1}^{\infty} |a^i|^p\right)^{1/p}, \quad 1 \le p < \infty. \tag{2.10}$$

Young's inequality says $|a^i b^i| \le |a^i|^p/p + |b^i|^q/q$ for $1/p + 1/q = 1$. (It suffices to take real positive a^i, b^i to prove it. Determine the maximum of the function $f_{b^i}(a^i) = \ b^i a^i - |a^i|^p/p$.) Therefore, if $1 < p < \infty, 1/p + 1/q = 1$, $||a||_p < \infty$, $||b||_q < \infty$ then $|\langle b, a\rangle| = |\sum b^i a^i| < \infty$, that is, $b \in l^q$ is a continuous linear functional on $l^p \ni a$, $l^q \subset l^{p*}$. It can be proved that $l^q = l^{p*}$ [2, Section IV.9]. Since X^* is always a Banach space, l^p, $1 < p < \infty$ is a Banach space. Additionally, the normed sequence spaces $l^\infty \supset c_0 \supset f$, all with norm

$$l^\infty: \ ||a||_\infty = \sup_i |a^i|, \tag{2.11}$$

$$c_0 \subset l^\infty: \ \lim_{i\to\infty} a^i = 0,$$

$$f \subset l^\infty: \ a^i = 0 \ \text{ for all but finitely many } i$$

are considered. It can be shown that l^∞ and c_0 are Banach spaces and $l^{1*} = l^\infty$ and $c_0^* = l^1$. Hence, l^1 is also a Banach space. It is easily seen that f has a countable base as a vector space. Moreover, it is dense in l^p, $1 \le p < \infty$ (in the topology of the norm $||\cdot||_p$) and in c_0 (in the topology of the norm $||\cdot||_\infty$). Hence, those spaces have a countable base and are separable. Finally, l^2 with the inner product $(a|b) = \sum_i \overline{a^i} b^i$ is the Hilbert space of **Heisenberg's quantum mechanics**. Every infinite-dimensional separable Hilbert space is isomorphic to l^2 [1, Section II.3].

$L^p(M, d\mu)$ [1]: Let $(M, d\mu)$ be a measure space, for instance $\mathbb{R}^n$ or a part of it with Lebesgue measure $d^n x$. Denote by f the class of complex or real functions on M which differ from each other at most on a set of measure zero. Clearly, linear combinations respect classes. $L^p(M, d\mu)$ is the functional linear space of classes f for which

$$||f||_p = \left(\int_M |f|^p d\mu\right)^{1/p} < \infty. \tag{2.12}$$

For $p = \infty$, $||f||_\infty = \text{ess sup}\,|f|$, that is the smallest real number c so that $|f| > c$ at most on a set of zero measure. For $1 \le p \le \infty$, $||f||_p$ is a norm, and $L^p(M, d\mu)$ is complete. $L^p(M, d\mu)^* = L^q(M, d\mu)$, $1/p + 1/q = 1$, $1 \le p < \infty$ with $\langle g, f\rangle = \int_M gfd\mu$. If $M = \mathbb{R}_+$ and $d\mu = \sum_{n=1}^\infty \delta(x-n)dx$, then $L^p(M, d\mu) = l^p$. If $\mu(M) < \infty$, then $L^p(M, d\mu) \subset L^{p'}(M, d\mu)$ for $p \ge p'$. The Hilbert space of Schrödinger's **quantum states** of a spinless particle is $L^2(\mathbb{R}^3, d^3x)$, for a spin-S particle is $L^2(\mathbb{R}^3, d^3x) \otimes \mathbb{C}^{2S+1}$, where $\mathbb{C}^{2S+1}$ is the $(2S+1)$-dimensional state space of spin. The L^p-spaces are for instance used in **density functional theories**.

Fock space: Let $\mathcal{H}$ be a Hilbert space of single-particle quantum states, and let $\mathcal{H}^0 = K$ (field of scalars) and $\mathcal{H}^n = \mathcal{H} \otimes \mathcal{H} \otimes \cdots \otimes \mathcal{H}$ (n factors). For any vector $\psi_{k_1} \otimes \psi_{k_2} \otimes \cdots \otimes \psi_{k_n} \in \mathcal{H}^n$, let $S_n\psi_{k_1} \otimes \psi_{k_2} \otimes \cdots \otimes \psi_{k_n} = \sum_{\mathcal{P}} \psi_{k_{\mathcal{P}(1)}} \otimes \psi_{k_{\mathcal{P}(2)}} \otimes \cdots \otimes \psi_{k_{\mathcal{P}(n)}}$ and $A_n\psi_{k_1} \otimes \psi_{k_2} \otimes \cdots \otimes \psi_{k_n} = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|} \psi_{k_{\mathcal{P}(1)}} \otimes \psi_{k_{\mathcal{P}(2)}} \otimes \cdots \otimes \psi_{k_{\mathcal{P}(n)}}$, where the summation is over all permutations $\mathcal{P}$ of the numbers $1, 2, \ldots, n$ and $|\mathcal{P}|$ is its order. Let $S_0 = A_0 = \text{Id}_{\mathcal{H}^0}$. Then,

$$\mathcal{F}_B(\mathcal{H}) = \oplus_{n=0}^\infty S_n \mathcal{H}^n$$

is the bosonic Fock space, and

$$\mathcal{F}_F(\mathcal{H}) = \oplus_{n=0}^\infty A_n \mathcal{H}^n$$

is the fermionic Fock space. An orthonormal base in both cases may be introduced as the set of occupation number eigenstates for a fixed orthonormal basis $\{\psi_k\}$ in $\mathcal{H}$

$|\rangle$, $|n_1, n_2, \ldots, n_N\rangle$, $N = 1, 2, \ldots$, $n_k = 0, 1, 2, \ldots$(bosons) and $n_k = 0, 1$ (fermions).

The state with vector $|\rangle \in \mathcal{H}^0$ is called the vacuum state. The Fock space is the closure (in the topology of the direct sum of tensor products of $\mathcal{H}$) of the span of all occupation number eigenstates.

2.3 Derivatives

Let $F : \Omega \to Y$ be a mapping (vector-valued function) from an open set Ω of a normed vector space X into a topological vector space Y. If the limes

$$D_xF(x_0) = \frac{d}{dt}F(x_0 + tx)|_{t=0} = \lim_{\substack{t \neq 0, t \to 0 \\ x_0 + tx \in \Omega}} \frac{F(x_0 + tx) - F(x_0)}{t} \tag{2.13}$$

exists it is called a partial derivative or (for $||x|| = 1$) **directional derivative** in the direction of x of the function F at x_0. $D_xF(x_0)$ is a vector of the space Y. $D_xF(x_0)$ is of course defined for any value of norm of x; by replacing in the above definition t by λt it is readily seen that $D_{\lambda x}F(x_0) = \lambda D_xF(x_0)$. (However, $D_xF(x_0)$ as a function of x need not be linear; for instance it may exist for some x and not for others.) If the directional derivative (for fixed x) exists for all $x_0 \in \Omega$ then $D_x F(x_0)$ is another function (of the variable x_0) from Ω into Y (which need not be continuous), and the second directional derivative $D_{x'} D_x F(x_0)$ may be considered if it exists for some x', and so on. If, given x_0, the directional derivative $D_x F(x_0)$ exists for all x as a continuous linear function from X into Y, then it is called the Gâteaux derivative.

Caution: The existence of all directional derivatives is not sufficient for the chain rule of differentiation to be valid; see example below.

Let Y also be a normed vector space. If there is a continuous linear function $DF(x_0) \in \mathcal{L}(X, Y)$ so that

$$F(x_0 + x) - F(x_0) = DF(x_0)x + R(x)||x||, \quad \lim_{x \to 0} R(x) = 0, \tag{2.14}$$

then $DF(x_0)$ is called the **total derivative** or the Fréchet derivative of F at x_0. $R(x)$ is supposed continuous at $x = 0$ with respect to the norm topologies of X and Y, and $R(0) = 0$. (For $x \neq 0$, $R(x)$ is uniquely defined to be $[F(x_0 + x) - F(x_0) - DF(x_0)x]/||x||$.) Given x (and x_0), $DF(x_0)x$ is again a vector in Y, that is, for given x_0, $DF(x_0)$ is a continuous linear function from X into Y. If $DF(x_0)$ exists for all $x_0 \in \Omega$, then DF is a mapping from Ω into $\mathcal{L}(X, Y)$ and DFx (x fixed) is a mapping from Ω into Y. Hence, the second derivative $D(DFx)(x_0)x' = D^2F(x_0)xx'$ may be considered, and so on. For instance, D^2F is a mapping from Ω into $\mathcal{L}(X, \mathcal{L}(X, Y))$, the space of continuous bilinear functions from $X \times X$ into Y and, given x and x', D^2Fxx' is a mapping from Ω into Y.

The total derivative may not exist even if all directional derivatives do exist. As an example [3, §10.1], consider $X = \mathbb{R}^2$, $Y = \mathbb{R}$ and the real function of two real variables x^1 and x^2

$$F(x^1, x^2) = \begin{cases} \dfrac{2(x^1)^3 x^2}{(x^1)^4 + (x^2)^2} & \text{for } (x^1, x^2) \neq (0, 0), \\ 0 & \text{for } (x^1, x^2) = (0, 0). \end{cases}$$

Let $0 = (0, 0)$ and $\boldsymbol{x} = (x^1, x^2) \neq 0$. Then, $(F(0 + t\boldsymbol{x}) - F(0))/t = (2t^3(x^1)^3 x^2)/(t^4(x^1)^4 + t^2(x^2)^2)$. For $x^2 = 0$ this is 0, and for $x^2 \neq 0$ it is of order $O(t)$, hence, $D_{\boldsymbol{x}}F(0) = 0$ for all $\boldsymbol{x}$. Nevertheless, $F(x^1, (x^1)^2) = x^1$: the slope of the graph of F on the curve $x^2 = (x^1)^2$ is unity. This means that $DF(0)$, which should be zero according to the directional derivatives, in fact does not exist: $R(\boldsymbol{x}) \to 0$ does not hold for $\boldsymbol{x} = (x^1, (x^1)^2)$. (Exercise: Show that $D_{\boldsymbol{x}}F(\boldsymbol{x}_0)$ is discontinuous at $\boldsymbol{x}_0 = 0$.)

If $D_x F(x_0')$ exists for all x and for all x_0' in a neighborhood U of x_0 and is continuous as a function of x_0' at x_0, then $DF(x_0)$ exists and $DF(x_0)x = D_x F(x_0)$.

Proof For small enough x so that $x_0 + x \in U$, consider the function $r(x_0, x) = F(x_0 + x) - F(x_0) - D_x F(x_0)$ with values in Y. Take any vector f of the dual space Y^* of Y and consider the scalar function $f(t) = \langle f, F(x_0 + tx) \rangle$ of the real variable $t, 0 \leq t \leq 1$. This function has a derivative

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \left\langle f, \frac{F(x_0 + tx + \Delta tx) - F(x_0 + tx)}{\Delta t} \right\rangle = \langle f, D_x F(x_0 + tx) \rangle$$

and hence $f(1) - f(0) = \langle f, D_x F(x_0 + \tau x) \rangle$ for some $\tau, 0 \le \tau \le 1$. Therefore, $\langle f, r(x_0, x) \rangle = \langle f, D_x F(x_0 + \tau x) - D_x F(x_0) \rangle$. Choose f with $||f|| = 1$ for which

$$|\langle f, r(x_0, x) \rangle| \ge \frac{1}{2} ||f|| \, ||r(x_0, x)|| = \frac{1}{2} ||r(x_0, x)||$$

holds. (It exists by the Hahn–Banach theorem.) It follows that $||r(x_0, x)|| \le 2|\langle f, D_x F(x_0 + \tau x) - D_x F(x_0) \rangle| \le 2||D_x F(x_0 + \tau x) - D_x F(x_0)||$. Finally, put $x = ||x|| \hat{x}$ and get $||r(x_0, x)|| \le 2||D_{\hat{x}}(x_0 + \tau x) - D_{\hat{x}}(x_0)|| \, ||x||$. Hence, in view of the continuity of $D_{\hat{x}}(x_0')$ at $x_0' = x_0$ it follows that $r(x_0, x) = R(x)||x||$ with $\lim_{x \to 0} R(x) = 0$. □

In the special case $Y = K$, the scalar field of X, the mapping $F : X \to K$ is a functional, and $DF(x_0) \in \mathcal{L}(X, K) = X^*$ is a continuous linear functional and hence an element of the dual space X^*, if it exists. For instance, if $X = K^n$ then $DF(x_0) = \boldsymbol{y} \in K^n$ (gradient). If X is a functional space, $DF(x_0)$ is called the **functional derivative** of F at x_0. If $X = L^p(K^n, d^n z) \ni f(z)$ then $DF(f_0) = g(z) \in L^q(K^n, d^n z), 1/p + 1/q = 1$. The functional derivative in the functional space L^p is a function (more precisely class of functions) of the functional space L^q. A trivial example which nevertheless is frequently met in physics is $F(f) = (g|f)$ with $D(g|f)(f) = g$ (derivative of a linear function).

If $X = K^n \ni \boldsymbol{x} = x^1 \boldsymbol{e}_1 + x^2 \boldsymbol{e}_2 + \cdots + x^n \boldsymbol{e}_n$ and $Y = K^m \ni \boldsymbol{y} = y^1 \boldsymbol{e}_1' + y^2 \boldsymbol{e}_2' + \cdots + y^m \boldsymbol{e}_m'$, then $\boldsymbol{F}(\boldsymbol{x}) = F^1(\boldsymbol{x}) \, \boldsymbol{e}_1' + F^2(\boldsymbol{x}) \, \boldsymbol{e}_2' + \cdots + F^m(\boldsymbol{x}) \boldsymbol{e}_m'$ and $\langle \boldsymbol{f}^i, \boldsymbol{DF}(\boldsymbol{x}_0) \boldsymbol{e}_k \rangle = \partial F^i(\boldsymbol{x}_0)/\partial x^k$, $\langle \boldsymbol{f}^i, \boldsymbol{e}_k' \rangle = \delta^i_k$. In this case,

$$\boldsymbol{DF}(\boldsymbol{x}_0) = \begin{pmatrix} \frac{\partial F^1(\boldsymbol{x}_0)}{\partial x^1} & \frac{\partial F^1(\boldsymbol{x}_0)}{\partial x^2} & \cdots & \frac{\partial F^1(\boldsymbol{x}_0)}{\partial x^n} \\ \frac{\partial F^2(\boldsymbol{x}_0)}{\partial x^1} & \frac{\partial F^2(\boldsymbol{x}_0)}{\partial x^2} & \cdots & \frac{\partial F^2(\boldsymbol{x}_0)}{\partial x^n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F^m(\boldsymbol{x}_0)}{\partial x^1} & \frac{\partial F^m(\boldsymbol{x}_0)}{\partial x^2} & \cdots & \frac{\partial F^m(\boldsymbol{x}_0)}{\partial x^n} \end{pmatrix} \tag{2.15}$$

is the **Jacobian matrix** of the function $\boldsymbol{F} : K^n \to K^m$. For any $\boldsymbol{y}^* \in Y^*$, $\langle \boldsymbol{y}^*, \boldsymbol{DF}(\boldsymbol{x}_0) \boldsymbol{x} \rangle = \boldsymbol{y}^* \cdot \boldsymbol{DF}(\boldsymbol{x}_0) \cdot \boldsymbol{x}$, where the dot $\cdot$ marks the inner product in the spaces K^n and K^m. For $m = n$ the determinant

$$\frac{D(y^1, \ldots, y^n)}{D(x^1, \ldots, x^n)} = \det\left(\frac{\partial F^i(\boldsymbol{x}_0)}{\partial x^j} \right) \tag{2.16}$$

is the **Jacobian**.

Employing higher derivatives, the **Taylor expansion** of a function F from the normed linear space X into a normed linear space Y reads

$$F(x_0 + x) = F(x_0) + DF(x_0)x + \frac{1}{2!} D^2 F(x_0) xx + \cdots + \frac{1}{k!} D^k F(x_0) \underbrace{xx \cdots x}_{(k \text{ factors})} + \cdots, \tag{2.17}$$

provided x_0 and $x_0 + x$ belong to a convex domain $\Omega \subset X$ on which F is defined and has total derivatives to all orders, which are continuous functions of x_0 in Ω and provided this Taylor series converges in the norm topology of Y. As explained after (2.14), $D^k F(x_0) \in \mathcal{L}(X, \mathcal{L}(X, \cdots, \mathcal{L}(X, Y) \cdots))$ is a k-linear function from $X \times X \times \cdots \times X$ (k factors) into Y. For instance, in the case $X = K^n, Y = K^m$ this means

$$\langle \boldsymbol{y}^*, D^k \boldsymbol{F}(\boldsymbol{x}_0)\boldsymbol{x} \cdots \boldsymbol{x}\rangle = \sum_{i,i_1,\ldots,i_k} y_i^* \frac{\partial F^i(\boldsymbol{x}_0)}{\partial x^{i_1} \cdots \partial x^{i_k}} x^{i_1} \cdots x^{i_k}. \tag{2.18}$$

Proofs of this Taylor expansion theorem and the following generalizations from standard analysis can be found in textbooks, for instance [4].

Recall that $\mathcal{L}(X, Y)$ is a normed vector space with the norm (2.5) which is Banach if Y is Banach. Hence, $\mathcal{L}(X, \mathcal{L}(X, Y))$ is again a normed vector space which is Banach if Y is Banach. If $L_2 : X \to \mathcal{L}(X, Y) : x, x' \mapsto L_2(x, x') =: L_2 x x'$ is a bilinear function from $X \times X$ into Y, its $\mathcal{L}(X, \mathcal{L}(X, Y))$-norm is (cf. (2.5))

$$\begin{aligned}||L_2||_{\mathcal{L}(X,\mathcal{L}(X,Y))} &= \sup_{x\in X} \frac{||L_2 x x'||_{\mathcal{L}(X,Y)}}{||x||_X} = \sup_{x\in X} \frac{\sup_{x'\in X} ||L_2 x x'||_Y / ||x'||_X}{||x||_X} \\ &= \sup_{x,x'\in X} \frac{||L_2 x x'||_Y}{||x||_X ||x'||_X}.\end{aligned}$$

By continuing this process, $\mathcal{L}(X, \mathcal{L}(X, \cdots, \mathcal{L}(X, Y) \cdots))$ (depth k) is a normed vector space which is Banach if Y is Banach, and the norm of a k-linear function $L_k x^{(1)} x^{(2)} \cdots x^{(k)}$ is

$$||L_k||_{\underbrace{\mathcal{L}(X, \mathcal{L}(X, \ldots, \mathcal{L}(X,Y)\ldots))}_{\text{depth } k}} = \sup_{x^{(1)}\ldots x^{(k)}\in X} \frac{||L_k x^{(1)} \cdots x^{(k)}||_Y}{||x^{(1)}||_X \cdots ||x^{(k)}||_X}. \tag{2.19}$$

A general **chain rule** holds for the case *if $F : X \supset \Omega \to Y, F(\Omega) \subset \Omega', G : Y \supset \Omega' \to Z$ and $H = G \circ F : X \supset \Omega \to Z$. Then,*

$$DH(x_0) = DG(F(x_0)) \circ DF(x_0) \tag{2.20}$$

if the right hand side derivatives exist. In this case, $DF(x_0) \in \mathcal{L}(X, Y)$ and $DG(F(x_0)) \in \mathcal{L}(Y, Z)$ and hence $DH(x_0) \in \mathcal{L}(X, Z)$. Moreover, if $DF : \Omega \to \mathcal{L}(X, Y)$ is continuous at $x_0 \in \Omega$ and $DG : \Omega' \to \mathcal{L}(Y, Z)$ is continuous at $F(x_0) \in \Omega'$, then $DH : \Omega \to \mathcal{L}(X, Z)$ is continuous at $x_0 \in \Omega$.

Coming back to the warning on p. 23, take the function $F : \mathbb{R} \to \mathbb{R}^2 : t \mapsto (t, t^2)$, and for $G : \mathbb{R}^2 \to \mathbb{R}$ take the function of the example on p. 23. Then, $H(t) = (G \circ F)(t) = t$ and hence $DH(0) = 1$. Would one from $D_{(x^1,x^2)}G(0,0) = 0$ for all (x^1, x^2) infer that $DG(0,0) = 0$, then one would get erroneously $DH(0) = DG(0,0) \circ DF(0) = 0$. In more familiar notation for this case,

$$\left.\frac{dH}{dt}\right|_{t=0} \neq \left.\frac{\partial G}{\partial x^1}\right|_{(0,0)} \left.\frac{dx^1}{dt}\right|_0 + \left.\frac{\partial G}{\partial x^2}\right|_{(0,0)} \left.\frac{dx^2}{dt}\right|_0 = 0.$$

The chain rule does not hold because the *total derivative* of G does not exist at $(0,0)$; $\partial G/\partial x^2$ is discontinuous there.

If X, Y and Z are the finite-dimensional vector spaces K^n, K^m and K^l with general (not necessarily orthonormal) bases fixed, then the $l \times n$ Jacobian matrix of $DH(x_0)$ is just the matrix product of the $l \times m$ and $m \times n$ Jacobian matrices (2.15) of $DG(F(x_0))$ and $DF(x_0)$. It follows that in the case $l = m = n$ the Jacobian of H is the product of the Jacobians of G and F:

$$\frac{D(z^1,\dots,z^n)}{D(x^1,\dots,x^n)} = \frac{D(z^1,\dots,z^n)}{D(y^1,\dots,y^n)}\frac{D(y^1,\dots,y^n)}{D(x^1,\dots,x^n)}.$$

Just this is suggested by the notation (2.16) of a Jacobian.

If $F : X \supset \Omega \to \Omega' \subset Y$ is a bijection and $DF(x_0)$ and $DF^{-1}(F(x_0))$ both exist, then

$$(DF(x_0))^{-1} = DF^{-1}(F(x_0)). \tag{2.21}$$

This follows from the chain rule in view of $F^{-1} \circ F = \mathrm{Id}_\Omega$ and $D\mathrm{Id}(x_0) = \mathrm{Id}$. (From the definition (2.14) it follows for a linear function $F \in \mathcal{L}(X,Y)$ that $DF(x_0) = F$ independent of $x_0 \in X$.) The case $X = Y = K^n$ now implies

$$\frac{D(x^1,\dots,x^n)}{D(y^1,\dots,y^n)} = \left(\frac{D(y^1,\dots,y^n)}{D(x^1,\dots,x^n)}\right)^{-1}$$

for the Jacobian. For $n = 1$ this is the rule $dx/dy = (dy/dx)^{-1}$.

A function F from an open domain Ω of a normed space X into a normed space Y is called a **class $C^n(\Omega, Y)$ function** if it has continuous derivatives $D^kF(x_0)$ up to order $k = n$ (continuous as functions of $x_0 \in \Omega$). If the domain Ω and the target space Y are clear from context, one speaks in short on a class C^n function (or even shorter of a C^n function). A C^0 function means just a continuous function. A C^∞ function is also called **smooth**. A smooth function still need not have a Taylor expansion. For instance the real function

$$f_\varepsilon(x) = \begin{cases} \exp(-\varepsilon^2/(\varepsilon^2 - x^2)) & \text{for } |x| < \varepsilon \\ 0 & \text{for } |x| \geq \varepsilon \end{cases}$$

is C^∞ on the whole real line, but has no Taylor expansion at the points $x = \pm\varepsilon$ although all its derivatives are equal to zero and continuous there. (Up to the normalization factor it is a δ_ε-function.) A function which has a Taylor expansion converging in the whole domain Ω is called a class $C^\omega(\Omega, Y)$ function or an **analytic function**. A complex-valued function of complex variables is analytic, iff it is C^1 and its derivatives obey the Cauchy–Riemann equations.

A C^n (C^{∞}, C^{ω}) **diffeomorphism** is a bijective mapping from $\Omega \subset X$ onto $\Omega' \subset Y$ which, along with its inverse, is $C^n, n > 0$, $(C^{\infty},\ C^{\omega})$.

With pointwise linear combinations of functions with constant coefficients, $(\lambda F + \lambda' G)(x) = \lambda F(x) + \lambda' G(x)$, the class C^n $(C^{\infty},\ C^{\omega})$ is made into a vector space. The vector spaces C^n, C^{∞} include normed subspaces C^n_b, C^{∞}_b (of all functions with finite norm) by introducing the norm

$$||F||_{C_b^{n/\infty}} = \sup_{\substack{x_0 \in \Omega \\ k \leq n/\infty}} ||D^k F(x_0)||, \quad ||F||_{C_b^0} = \sup_{x_0 \in \Omega} ||F(x_0)||, \tag{2.22}$$

with the norms (2.19) on the right hand side of the first expression. These spaces are again Banach if Y is Banach. Convergence of a sequence of functions in these norms means uniform convergence on Ω, of the sequence of functions and of the sequences of all derivatives up to order n, or of unlimited order. (Besides, every space $C^n_b, m \leq n \leq \infty$, is dense in the normed space C^m_b.)

The mapping $D : C^1_b(\Omega, Y) \to C^0_b(\Omega, \mathcal{L}(X, Y)) : F \mapsto DF$ *is a continuous linear mapping with norm not exceeding unity.*

Proof As a bounded linear mapping, $D \in \mathcal{L}(C^1_b(\Omega, Y), C^0_b(\Omega, \mathcal{L}(X, Y)))$, the norm of D is $||D|| = \sup_F ||DF||_{C^0_b(\Omega, \mathcal{L}(X,Y))} / ||F||_{C^1_b(\Omega, Y)}$. From (2.22) it is directly seen that the numerator of this quotient cannot exceed the denominator, hence $||D|| \leq 1$ and D is indeed bounded and hence continuous. □

If the normed vector space Y in addition is an algebra with unity I (see Compendium) and the norm has the additional properties

4. $||I|| = 1$,
5. $||yy'|| \leq ||y||\,||y'||$,

then it is called a **normed algebra**. If it is complete as a normed vector space, it is called a **Banach algebra**. If Y is a normed algebra, then with pointwise multiplication, $(FG)(x) = F(x)G(x)$, the class C^n_b (C^{∞}_b) with the norm (2.22) is made into a normed algebra. (Show that FG is C^n_b if F and G both are C^n_b.)

The **derivative of a product** in the algebra $C^n, n \geq 1$ is obtained by the **Leibniz rule**

$$D(FG) = (DF)G + F(DG). \tag{2.23}$$

(Exercise: Consider $\Phi(x) = (F(x), G(x))$, $\Psi(u, v) = uv$ and $H(x) = (\Psi \circ \Phi)(x)$ and apply the chain rule to obtain (2.23).)

An **implicit function** is defined in general in the following manner: *Let X be a topological space, let Y be a Banach space and let Z be a normed vector space. Let $F : X \times Y \supset \Omega \to Z$ be a continuous function and consider the equation*

$$F(x, y) = c, \quad c \in Z \text{ fixed}. \tag{2.24}$$

Assume that $D_yF(x_0, y_0) \in \mathcal{L}(Y, Z)$ *exists for all* $y \in Y$ *and is continuous on* Ω *(as a function of* x_0, y_0*), that* $F(a, b) = c$ *and that* $Q = D_yF(a, b)$ *is a linear bijection from* Y *onto* Z*, so that* $Q^{-1} \in \mathcal{L}(Z, Y)$*. Then, there are open sets* $A \ni a$ *and* $B \ni b$ *in* X *and* Y*, so that for every* $x \in A$ Eq. (2.24) *has a unique solution* $y \in B$ *which implicitly by* Eq. (2.24) *defines a continuous function* $G : A \to Y : x \mapsto y = G(x)$.

The proofs of this theorem and of the related theorems below are found in textbooks, for instance [4]. It is essential, that Y is Banach.

Let X *be also a normed vector space and assume* $F \in C^1(\Omega, Z)$*. Then the above function* G *has a continuous total derivative at* $x = a$*, and*

$$D_xG(a) = -(D_yF(a,b))^{-1} \circ D_xF(a,b), \quad b = G(a). \tag{2.25}$$

Formally, one may differentiate (2.24) by applying the chain rule,

$$D_xF(a,b)dx + D_yF(a,b)dy = 0, \quad x \in X,\ y \in Y,$$

where $dx = D\mathrm{Id}_X = 1$ and $dy = D_xG(a)$, and solve this relation for dy/dx.

In order to prove that $DG(a)$ of (2.25) is a continuous function of a, that is, that $G \in C^1(A, Y)$, the continuity of $(D_yF(a,b))^{-1}$ as function of a and b must be stated. Since $D_yF(a,b) \in \mathcal{L}(Y, Z)$, this implies the derivative of the inverse of a linear function with respect to a parameter which is of interest on its own:

Let X *and* Y *be Banach spaces and let* $\mathcal{U}$ *and* $\mathcal{U}^{-1}$ *be the sets of invertible continuous linear mappings out of* $\mathcal{L}(X, Y)$ *and* $\mathcal{L}(Y, X)$*. Then, both* $\mathcal{U}$ *and* $\mathcal{U}^{-1}$ *are open sets.*

Proof for $\mathcal{U}$*; for* $\mathcal{U}^{-1}$ *interchange* X *and* Y Let $U_0 \in \mathcal{U}$ and $U \in \mathcal{L}(X, Y)$ such that $||\mathrm{Id}_X - U_0^{-1} \circ U||_{\mathcal{L}(X,X)} < 1$. Then,

$$(U_0^{-1} \circ U)^{-1} = \mathrm{Id}_X + (\mathrm{Id}_X - U_0^{-1} \circ U) + (\mathrm{Id}_X - U_0^{-1} \circ U)^2 + \cdots$$

converges and hence $U = U_0 \circ (U_0^{-1} \circ U) \in \mathcal{U}$ $(U^{-1} = (U_0^{-1} \circ U)^{-1} \circ U_0^{-1})$. Every $U_0 \in \mathcal{U}$ has a neighborhood, $||U_0 - U|| < 1/||U_0||$, where this is realized. □

Let X *and* Y *be Banach spaces and* $\mathcal{U}$ *as above. Let* $\Phi : X \supset A \to \mathcal{U} \subset \mathcal{L}(X, Y) : x_0 \mapsto U(x_0)$ *be* C^1*. Then* $\tilde{\Phi} : x_0 \mapsto (U(x_0))^{-1}$ *is* C^1*, and its derivative is given by*

$$D(\tilde{\Phi})(x_0)x = -\tilde{\Phi}(x_0) \circ D\Phi(x_0)x \circ \tilde{\Phi}(x_0) \in \mathcal{L}(Y, X), \quad x \in X. \tag{2.26}$$

The proof of continuity of the left hand side with respect to the x_0-dependence consists of an investigation of the relation $\Phi(x) \circ \tilde{\Phi}(x) = \mathrm{Id}_Y$. It is left to the reader (see textbooks of analysis). Differentiating this equation with respect to x at point x_0 yields $\Phi(x_0) \circ D\tilde{\Phi}(x_0)x + D\Phi(x_0)x \circ \tilde{\Phi}(x_0) = 0$. Composing with $(\Phi(x_0))^{-1} = \tilde{\Phi}(x_0)$ from the left results in the above relation. If $X = K^n$, then $\mathcal{U}$ can only be non-empty if also $Y = K^n$. After introducing bases $U(\boldsymbol{x})$ is represented by a regular $n \times n$ matrix $\boldsymbol{M}(\boldsymbol{x})$. One obtains the familiar result $(\boldsymbol{x} \cdot \partial/\partial\boldsymbol{x})\boldsymbol{M}^{-1}|_{x_0} = \boldsymbol{M}^{-1} \cdot (\boldsymbol{x} \cdot \partial/\partial\boldsymbol{x})\boldsymbol{M}|_{x_0} \cdot \boldsymbol{M}^{-1}$. Along a straight line $\boldsymbol{x} = t\boldsymbol{e}$, or for a one parameter dependent matrix this reduces to $d\boldsymbol{M}^{-1}/dt = \boldsymbol{M}^{-1} \cdot (d\boldsymbol{M}/dt) \cdot \boldsymbol{M}^{-1}$.

2.4 Compactness

Compactness is the abstraction from closed bounded subsets of $\mathbb{R}^n$. Before introducing this concept, a few important properties of n-dimensional closed bounded sets are reviewed.

The **Bolzano–Weierstrass theorem** says that in an n-dimensional closed bounded set every sequence has a convergent subsequence. An equivalent formulation is that every infinite set of points of an n-dimensional closed bounded set has a cluster point.

A **cluster point** of a subset A of a topological space X is a point $x \in X$ every neighborhood of which contains at least one point of *A distinct from x*. (Compare the definition of a point of closure on p. 12. A cluster point is a point of closure, but the reverse is not true in general.)

Weierstrass theorem: A continuous function takes on its maximum and minimum values on an n-dimensional closed bounded set.

Brouwer's fixed point theorem: On a *convex* n-dimensional closed bounded set B the fixed point equation $x = F(x), F : B \to B$ continuous, has a solution.

These theorems do not necessarily hold in infinite dimensional spaces. Consider for example the closed unit ball (e.g. centered at the origin) in an infinite dimensional real Hilbert space. Clearly the sequence of distinct orthonormal unit vectors does not converge in the norm topology: the distance between any pair of orthogonal unit vectors is $||e_i - e_j|| = (e_i - e_j|e_i - e_j)^{1/2} = \sqrt{2}$. It is easily seen that open balls of radius $1/(2\sqrt{2})$ centered halfway on these unit vectors do not intersect. The unit ball is too roomy for the Bolzano–Weierstrass theorem to hold; it accommodates an infinite number of non-overlapping balls of a fixed non-zero radius. This consideration yields the key to compactness.

A set C of a topological space is called a **compact set**, if every open cover $\{U\}$, a family of open sets with $\cup U \supset C$, contains a finite subcover, $\cup_{i=1}^{n} U_i \supset C$. A compact set in a Hausdorff space (the only case of interest in this volume) is called a **compactum**.

Compactness is a topological property, *the image C' of a compact set C under a continuous mapping F is obviously a compact set*: Take any open cover of C'. Since the preimage $F^{-1}(U')$ of an open set U' is an open set $U \subset C$, these preimages form an open cover of C. A selection of a finite subcover of these preimages also selects a finite subcover of C'.

A compactum is closed.

Proof Let x be a point of closure of a compactum C, that is, every neighborhood of x contains at least one point $c \in C$. Let $x \notin C$. Since C is Hausdorff, for every $c \in C$ there are disjoint open sets $U_c \ni c$ and $V_{x,c} \ni x$. Since the sets U_c obviously form an open cover of C, a finite subcover $U_{c_i}, i = 1, \ldots, n$ may be selected. Then, $V = \cap_i V_{x,c_i}$ is a neighborhood of x not intersecting C, which contradicts the preposition. Hence, C contains all its points of closure. □

It easily follows that *the inverse of a continuous bijection f of a compactum C onto a compactum C' is continuous, that is, the bijection is a homeomorphism.*

Proof Indeed, any closed subset A of the compactum C is a compactum; any open cover of A together with the complement of A forms an open cover of C and hence there is a finite subcover which is also a subcover of A. Now, since f is continuous, $f(A)$ is also a compactum and hence a closed subset of C'. Consequently f maps closed sets to closed sets, and because it is a bijection, it also maps open sets to open sets. □

Now, the Bolzano–Weierstrass theorem is extended:

Every infinite set of points of a compact set C has a cluster point.

Proof Assume that the infinite set $A \subset C$ has no cluster point. A set having no cluster point is closed. Indeed, if a is a point of closure of A, then $a \in A$ or a is a cluster point of A. Select any infinite sequence $\{a_i\} \subset A$ of distinct points a_i. The sets $\{a_i\}_{i=n}^{\infty}$ are closed for $n = 1, 2, \ldots$ and the intersection of any finite number of them is not empty. Their complements U_n in C form an open cover of C, for which hence there exists no finite subcover. C is not a compact set. □

As a consequence, an unbounded set of a metric space cannot be compact. Hence, the simple Heine–Borel theorem, that a closed bounded subset of $\mathbb{R}^n$, $n < \infty$ is compact, has a reversal: A compact subset of $\mathbb{R}^n$ is closed and bounded. (Recall that a metric space is Hausdorff.) This immediately also extents the Weierstrass theorem:

A continuous real-valued function on a compact set takes on its maximum and minimum values.

It maps the compact domain onto a compact set of the real line, which is closed and bounded and hence contains its minimum and maximum. However, a much more general statement on the existence of extrema will be made later on.

A closed subset of a compact set is a compact set.

Proof Take any open cover of the closed subset C' of the compact set C. Together with the set $C \backslash C'$, open in C, it also forms an open cover of C. A finite subcover of C also yields a finite subcover of C'. □

A set of a topological space is called **relatively compact** if its closure is compact. A topological space is called **locally compact** if every point has a relatively compact neighborhood. A function from a domain in a metric space X into a metric space Y is called a **compact function** or compact operator if it is continuous and maps bounded sets to relatively compact sets.

Brouwer's fixed point theorem has now two important generalizations which are given without proof (see textbooks of functional analysis):

Tychonoff's fixed point theorem: *A continuous mapping $F : C \rightarrow C$ in a compact convex set C of a locally convex vector space has a fixed point.*

Schauder's fixed point theorem: *A compact mapping $F : C \rightarrow C$ in a closed bounded convex set C of a Banach space has a fixed point.*

Both theorems release the precondition of Banach's fixed point theorem on F to be a strict contraction (p. 14). As a price, uniqueness is not guaranteed any more. Tychonoff's theorem also releases the precondition of completeness of the space.

Every locally compact space has a ***one point compactification*** *that is, a compact space* $X^c = X \cup \{x_\infty\}$ *and a homeomorphism* $P : X \to X^c \setminus \{x_\infty\}$.

Proof Let $x_\infty \notin X$ and let $\{U\}_\infty$ be the class of open sets of X for which $X \setminus U$ is compact in X. (X itself belongs to this class since $\emptyset$ is compact.) Take the open sets of X^c to be the open sets of X and all sets containing x_∞ and having their intersections with X in $\{U\}_\infty$. This establishes a topology in X^c and the homeomorphism. Let now $\{V\}$ be an open cover of X^c. It contains at least one set $V_\infty = U \cup \{x_\infty\}$, and $X^c \setminus V_\infty$ is compact in X. Hence, $\{V\}$ has a finite subcover. □

The compactified real line (circle) $\overline{\mathbb{R}}$ and the compactified complex plane (Riemann sphere) $\overline{\mathbb{C}}$ are well known examples of one point compactifications.

To get more general results for the existence of extrema, the concept of semicontinuity is needed. A function F from a domain of a topological space X into $\overline{\mathbb{R}}$ is called **lower (upper) semicontinuous** at the point $x_0 \in X$, if either $F(x_0) = -\infty$ ($F(x_0) = +\infty$) or for every $\varepsilon > 0$ there is a neighborhood of x_0 in which $F(x) > F(x_0) - \varepsilon$ ($F(x) < F(x_0) + \varepsilon$).

A lower semicontinuous function need not be continuous, its function value even may jump from $-\infty$ to ∞ at points of discontinuity. However, at every point of discontinuity it takes on the lowest limes of values. (For every net converging towards $x_0 \in X$ the function value at x_0 is equal to the lowest cluster point of function values on the net.) A lower semicontinuous function is **finite from below**, if $F(x) > -\infty$ for all x. Analogous statements hold for an upper semicontinuous function.

If F is a finite from below and lower semicontinuous function from a non-empty compactum A into $\mathbb{R}$, *then F is even bounded below and the minimum problem* $\min_{x\in A} F(x) = \alpha$ *has a solution* $x_0 \in A, \alpha = F(x_0)$.

An analogous theorem holds for a maximum problem. The proof of these statements is simple: Consider the infimum of F on A, pick a sequence for which $F(x_n) \leq \inf F(x) + 1/n$ and select a cluster point x_0 and a subnet converging to x_0. Hence, $\inf F(x) = F(x_0) > -\infty$ since F is finite from below.

Extremum problems are ubiquitous in physics. Many physical principles are directly variational. Extremum problems are also in the heart of duality theory which in physics mainly appears as theory of Legendre transforms. Moreover, since every system of partial differential equations is equivalent to a variational problem, extremum problems are also central in (particularly non-linear) analysis, again with central relevance for physics.

It has become evident above that compactness of the domain plays a decisive role in extremum problems. On the other hand, bounded sets in infinite-dimensional normed spaces are not compact in the norm topology, while many variational problems, in particular in physics, are based on infinite-dimensional functional spaces. (David Hilbert introduced the concept of functional inner product space to bring forward the variational calculus.) Rephrased, those functional spaces are not locally compact in the norm topology. The question arises, can one introduce a more cooperative topology in those spaces. The coarser a topology, the less open sets exist, and the more chances appear for a set to be compact. On p. 18, the weak topology was introduced as the coarsest topology in the vector space X, for which all bounded linear functionals are continuous. In a finite-dimensional space it was shown to be equivalent to the norm topology. In an infinite-dimensional space it is indeed coarser than the norm topology, but sometimes not coarse enough to our goal.

Let X be a Banach space and X^* its dual. In general, X^{**}, the space of all bounded linear functionals on X^*, may be larger than X. The weak topology of X^* is the coarsest topology in which all bounded linear functionals, that is all $f \in X^{**}$ are continuous. The **weak* topology** is the coarsest topology of X^* in which all bounded functionals $f \in X$ are continuous. Since these are in general less functionals, in general the weak* topology is coarser than the weak topology. If X is reflexive, then $X^{**} = X$ (and $X^{***} = X^*$), and the weak and weak* topologies of X^* (and also of $X^{**} = X$) are equivalent. (A Banach space is in general not any more first countable in the weak and weak* topologies; this is why instead of sequences nets are needed.)

The Banach–Alaoglu theorem states *that the unit ball of the dual X^* of a Banach space X is compact in the weak* topology*. As a corollary, *the unit ball of a reflexive Banach space is compact in the weak topology*.

A proof which uses Tichonoff's non-trivial theorem on topological products may be found in textbooks on functional analysis. Now, the way is paved for applications of the existence theorems of extrema. The price is that in the weak* topology there are much less semicontinuous functions than in the norm topology. Nevertheless, for instance the theory of functional Legendre transforms, relevant in **density functional theories** is pushed far ahead [5, and citations therein].

A few applications of the concept of compactness in functional analysis are finally mentioned which are related to the material of this volume. They use the facts that every compactum X is a **regular topological space**, that is, every non-empty open set contains the closure of another non-empty open set, and every compactum is a **normal topological space**, which means that every single point set $\{x\}$ is closed and every pair of disjoint closed sets is each contained in one of a pair of disjoint open sets.

Proof For each pair (x, y) of points in a pair of disjoint closed sets (C_1, C_2), $C_1 \cap C_2 = \emptyset, x \in C_1, y \in C_2$, there is a pair of disjoint open sets $(U_{x,y}, U_{y,x})$, $x \in U_{x,y}$, $y \in U_{y,x}$, since a compactum is Hausdorff. C_2 as a closed subset of a compactum is compact, and hence has a finite open cover $\{U_{y_1,x}, U_{y_2,x}, \ldots, U_{y_n,x}\}$, $y_i \in C_2$. Put $U_x = \cup_i U_{y_i,x}, U^x = \cap_i U_{x,y_i}$, and $U_1 = \cup_j U_{x_j}, U_2 = \cap_j U^{x_j}$, $x_j \in C_1$.

Regularity: Let U be an open set. Put $C_1 = X \backslash U$ and $C_2 = \{y\}$, $y \in U$ (C_2 is closed since X is Hausdorff.). Take $U_2 \ni y$ constructed above. $\overline{U_2} \subset U$.

Normality: For the above constructions, obviously $C_1 \subset U_1$, $C_2 \subset U_2$, $U_1 \cap U_2 = \emptyset$. $\square$

In a regular topological space every point has a *closed neighborhood base*.

For the proofs of the following theorems see textbooks of functional analysis.

Urysohn's theorem: *For every pair* (C_0, C_1) *of disjoint closed sets of a normal space* X *there is a real-valued continuous function,* $F \in C^0(X, \mathbb{R})$, *with the properties* $0 \le F(x) \le 1, F(x) = 0$ *for* $x \in C_0$, $F(x) = 1$ *for* $x \in C_1$.

Tietze's extension theorem: *Let* X *be a compactum and* $C \subset X$ *closed. Then every* $C^0(C, \mathbb{R})$*-function has a* $C^0(X, \mathbb{R})$*-extension.*

A function F defined on a locally compact topological space X with values in a normed vector space Y is said to be a **function of compact support**, if it vanishes outside of some compact set (in general depending on F). The **support of a function** $F, \operatorname{supp} F$ is the smallest closed set outside of which $F(x) = 0$. If X is a locally compact normed vector space, then corresponding to the classes $C^n, 0 \le n \le \infty$ (p. 27) there are **classes** C_0^n of continuous or n times continuously differentiable functions of compact support. Like the classes C^n, the classes C_0^n are vector spaces or in the case of an algebra Y algebras with respect to pointwise operations on functions.

In the context of this volume, particularly $C_0^\infty(K^n, Y)$ functions, $K = \mathbb{R}$ or $\mathbb{C}$, are of importance. One could normalize the vector space $C_0^n, 0 \le n \le \infty$ with the C_b^n-norm (2.22), however, if X itself is not compact, C_{0b}^n would not be complete in this norm topology even if Y would be Banach. For instance, the function sequence

$$F_n(x) = \sum_{k=1}^{n} \frac{1}{2^k} \Phi(x-k), \quad \Phi \in C_0^n(\mathbb{R}, \mathbb{R}), \ \ n = 1, 2, \dots,$$

is Cauchy in the C_b^n norm, but its limit does not have compact support. The completion of the $C_0^0(X, Y) = C_{0b}^0(X, Y)$ space of continuous functions of compact support in the C_b^0-norm is the space $C_\infty(X, Y)$ of continuous functions vanishing for $||x||_X \to \infty$, that is, for every $\varepsilon > 0$ there is a compact $C_\varepsilon \subset X$ outside of which $||F(x)||_Y < \varepsilon$. (Hence, $C_0^0(X, Y)$ is dense in $C_\infty(X, Y)$ in the C_b^0-norm; moreover, all $C_0^n(X, Y)$, $0 \le n \le \infty$ are dense in $C_\infty(X, Y)$ in the C_b^0-norm. If X is not compact, of course non of those classes is dense in any C_b^n in the C_b^n-norm: Let for instance $||F_1(x)||_Y = 1$ for all $x \in X$, then $||F - F_1||_{C_b^n} = 1$ for all $F \in C_0^m$. These are simple statements on uniform approximations of functions by more well behaved functions.)

Functions of compact support are very helpful in analysis, geometry and physics. They are fairly wieldy since their study is much the same as that of functions on a closed bounded subset of $\mathbb{R}^n$. The tool of continuation of structures from this rather simple situation to much more complex spaces, that is to connect local with global structures, is called partition of unity. It works for all locally compact spaces which are countable unions of compacta. (Caution: Not every countable union of compact sets is locally compact.) However, the most general class of spaces where it works are the paracompact spaces.

A **paracompact space** is a Hausdorff topological space for which every open cover, $X \subset \cup_{\alpha\in A} U_\alpha$, has a **locally finite** refinement, that is an open cover $\cup_{\beta\in B} V_\beta$ for which every V_β is a subset of some U_α and every point $x \in X$ has a neighborhood W_x which intersects with a finite number of sets V_β only.

A **partition of unity** on a topological space X is a family $\{\varphi_\alpha | \alpha \in A\}$ of $C_0^\infty(X, \mathbb{R})$-functions such that

1. there is a locally finite open cover, $X \subset \cup_{\beta\in B} U_\beta$,
2. the support of each φ_α is in some U_β, $\{\operatorname{supp} \varphi_\alpha | \alpha \in A\}$ is locally finite,
3. $0 \le \varphi_\alpha(x) \le 1$ on X for every α,
4. $\sum_{\alpha\in A} \varphi_\alpha(x) = 1$ on X.

The last sum is well defined since, given x, only a finite number of items are non-zero due to the locally finite cover governing the partition. The partition of unity is called **subordinate** to the cover $\cup_{\beta\in B} U_\beta$.

A paracompact space could also be characterized as a space which permits a partition of unity. It can be shown that every second countable locally compact Hausdorff space is paracompact. This includes locally compact Hausdorff spaces which are countable unions of compact sets, in particular it includes $\mathbb{R}^n$ for finite n. However, *any* (not necessarily countable) *disconnected* union (see next section) of paracompact spaces is also paracompact.

The function f_ε on p. 26 is an example of a real C_0^∞-function on $\mathbb{R}$. A simple example of functions φ_α on $\mathbb{R}^n$ is obtained by starting with the C^∞-function

$$f(t) = \begin{cases} e^{-1/t} & \text{for } t > 0 \\ 0 & \text{for } t \le 0 \end{cases}$$

and putting $g(t) = f(t)/(f(t) + f(1-t))$, which is C^∞, $0 \le g(t) \le 1, g(t) = 0$ for $t \le 0, g(t) = 1$ for $t \ge 1$. Then, $h(t) = g(t+2)g(2-t)$ is C_0^∞, $0 \le h(t) \le 1, h(t) = 0$ for $|t| \ge 2, h(t) = 1$ for $|t| \le 1$. Now, with a dual base $\{\boldsymbol{f}^i\}$ in $\mathbb{R}^n$, the $C_0^\infty(\mathbb{R}^n, \mathbb{R})$-function

$$\psi(\boldsymbol{x}) = h(x^1)h(x^2)\cdots h(x^n), \quad x^i = \boldsymbol{f}^i \cdot \boldsymbol{x}, \tag{2.27}$$

has the properties $0 \le \psi(\boldsymbol{x}) \le 1, \psi(\boldsymbol{x}) = 0$ outside the compact n-cube $|x^i| \le 2$ with edge length 4 which is contained in an open n-cube with edge length $4 + \varepsilon$, $\varepsilon > 0$, and $\psi(\boldsymbol{x}) = 1$ inside the n-cube with edge length 2, all centered at the origin of $\mathbb{R}^n$. The total $\mathbb{R}^n$ may be covered with open n-cubes of edge length $4 + \varepsilon$ centered at points $\boldsymbol{m} = (x^1 = 3m_1, x^2 = 3m_2, \ldots, x^n = 3m_n)$, m_i integer. Then,

$$\varphi_{\boldsymbol{m}}(\boldsymbol{x}) = \frac{\psi(\boldsymbol{x} - \boldsymbol{m})}{\sum_{\boldsymbol{m}'} \psi(\boldsymbol{x} - \boldsymbol{m}')}, \quad \sum_{\boldsymbol{m}} \varphi_{\boldsymbol{m}}(\boldsymbol{x}) = 1, \tag{2.28}$$

is a partition of unity on $\mathbb{R}^n$ (Fig. 2.5).

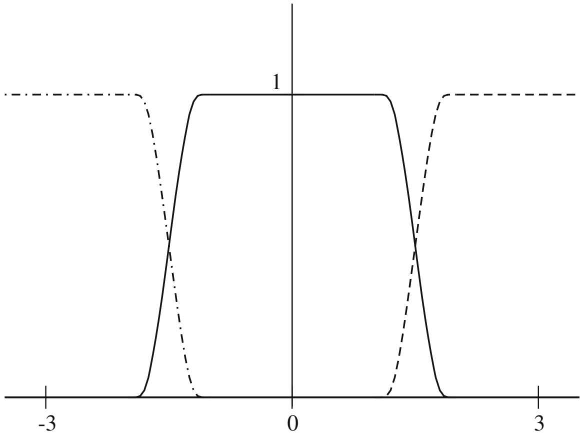

Fig. 2.5 Partition of unity on $\mathbb{R}$ with functions (2.28)

Besides applications in the theory of generalized functions and in the theory of manifolds, the partition of unity has direct applications in physics. For instance in **molecular orbital theory** of molecular or solid state physics the single particle quantum state (molecular orbital) is expanded into local basis orbitals centered at atom positions. For convenience of calculations one would like to have the density and self-consistent potential also as a site expansion of local contributions, hopefully to be left with a small number of multi-center integrals. This is however not automatically provided since the density is bilinear in the molecular orbitals, and the self-consistent potential is non-linear in the total density. If $v(\boldsymbol{x})$ is the self-consistent potential in the whole space $\mathbb{R}^3$ and $\sum_{\boldsymbol{R}} \varphi_{\boldsymbol{R}}(\boldsymbol{x})$ is a partition of unity on $\mathbb{R}^3$ with functions centered at the atom positions $\boldsymbol{R}$, then

$$v(\boldsymbol{x}) = \sum_{\boldsymbol{R}} (v(\boldsymbol{x})\varphi_{\boldsymbol{R}}(\boldsymbol{x})) = \sum_{\boldsymbol{R}} v_{\boldsymbol{R}}(\boldsymbol{x})$$

is the wanted expansion with potential contributions $v_{\boldsymbol{R}}$ of compact support. Thus, the number of multi-center integrals can be made finite in a very controlled way.

Finally, distributions (generalized functions) with compact support are shortly considered which comprise Dirac's δ-function and its derivatives.

Consider the whole vector space $C^\infty(\mathbb{R}^n, \mathbb{R})$ and instead of (2.22) for every compact $C \subset \mathbb{R}^n$ introduce the seminorm

$$p_{C,m}(F) = \sup_{\substack{\boldsymbol{x}\in C\\ |l|\le m}} |D^l F(\boldsymbol{x})|, \quad D^0 F = F, \quad l = (l_1, \dots, l_n),\ l_j \ge 0,$$
$$D^l F(\boldsymbol{x}) = \frac{\partial^{l_1+l_2+\cdots+l_n}}{(\partial x^1)^{l_1}(\partial x^2)^{l_2}\cdots(\partial x^n)^{l_n}} F(x^1, x^2, \dots, x^n), \quad l_1 + l_2 + \cdots + l_n = |l|. \tag{2.29}$$

It is a seminorm because it may be $p_{C,m}(F) = 0, F \neq 0$ (if $\operatorname{supp} F \cap C = \emptyset$). In the topology of the family of seminorms for all $C \subset \mathbb{R}^n$ and all m, convergence of a sequence of functions means uniform convergence of the functions and of all their derivatives on every compactum. This topology is a metric topology, and the vector space $C^\infty(\mathbb{R}^n, \mathbb{R})$ topologized in this way is also denoted $\mathcal{E}(\mathbb{R}^n, \mathbb{R})$ or in short $\mathcal{E}$.

Indeed, *consider a sequence of compacta* $C_1 \subset C_2 \subset \cdots$ *with* $\cup_{i=1}^\infty C_i = \mathbb{R}^n$ *(for instance closed balls with a diverging sequence of radii). Then, the function*

$$d(F,G) = \sum_{i=1}^\infty 2^{-i} \frac{d_{C_i}(F,G)}{1 + d_{C_i}(F,G)}, \quad d_C(F,G) = \sum_{m=0}^\infty 2^{-m} \frac{p_{C,m}(F-G)}{1 + p_{C,m}(F-G)} \tag{2.30}$$

is a distance function.

Proof Clearly, $d(F,G) \neq 0$, if $F(\boldsymbol{x}) \neq G(\boldsymbol{x})$ for some $\boldsymbol{x}$ since the C_i cover $\mathbb{R}^n$. To prove the triangle inequality, consider the obvious inequality $(\alpha+\beta)/(1+\alpha+\beta) \leq \alpha/(1+\alpha) + \beta/(1+\beta)$ for any pair α, β of non-negative real numbers. In view of $|\alpha-\beta| \leq |\alpha-\gamma| + |\gamma-\beta|$ for any three real numbers α, β, γ it follows $|\alpha-\beta|/(1+|\alpha-\beta|) \leq |\alpha-\gamma|/(1+|\alpha-\gamma|) + |\gamma-\beta|/(1+|\gamma-\beta|)$. This yields the triangle inequality for each fraction on the right hand side of the second equation (2.30). Since each of these fractions is ≤ 1, the series converges to a finite number also obeying the inequality for d_C. For d it is obtained along the same line. □

Any topological vector space the topology of which is given by a *countable, separating* family of seminorms, which means that the difference of two distinct vectors has at least one non-zero seminorm, can be metrized in the above manner.

$\mathcal{E}(\mathbb{R}^n, \mathbb{R})$ *is a Fréchet space.*

Proof Completeness has to be proved. In $\mathcal{E}$, $\lim_{i,j\to\infty} d(F_i, F_j) = 0$ means that on every compactum $C \subset \mathbb{R}^n$ the sequence F_i together with the sequences of all derivatives converge uniformly. Hence, on every C and consequently on $\mathbb{R}^n$ the limit exists and is a C^∞-function F. □

The elements f of the dual space $\mathcal{E}^*$ of $\mathcal{E}$, that is the bounded linear functionals on $\mathcal{E}$, are called **distributions** or **generalized functions**. $\mathcal{E}$ is called the **base space** of the distributions $f \in \mathcal{E}^*$. Formally, the writing

$$\langle f, F\rangle = \int d^n x f(\boldsymbol{x}) F(\boldsymbol{x}), \quad F \in \mathcal{E}, \tag{2.31}$$

is used based on the linearity in F of integration. However, $f(\boldsymbol{x})$ has a definite meaning only in connection with this integral. Every ordinary L^1-function f with compact support defines via the integral (2.31) in the Lebesgue sense a bounded linear functional on $\mathcal{E}$, hence these functions (more precisely, equivalence classes of functions forming the elements of L^1) are special $\mathcal{E}^*$-distributions. Derivatives of distributions are defined via the derivatives of functions $F \in \mathcal{E}$ by formally integrating by parts. Hence, per definition distributions have derivatives to all orders. This holds also for L^1-functions (with compact support) considered as distributions. Derivatives of discontinuous functions as distributions comprise **Dirac's δ-function**

$$\langle \delta_{\boldsymbol{x}_0}, F\rangle = \int d^n x\, \delta(\boldsymbol{x} - \boldsymbol{x}_0) F(\boldsymbol{x}) = F(\boldsymbol{x}_0).$$

Elements of $\mathcal{E}^*$ are not the most general distributions. In the spirit of formula (2.31), more general distributions are obtained by narrowing the base space. In physics, **densities and spectral densities** are in general distributions, if they comprise point masses or point charges or point spectra (that is, eigenvalues).

Let $U \subset \mathbb{R}^n$ be open and consider all $F \in \mathcal{E}$ with $\operatorname{supp} F \subset U$. If $\langle f, F\rangle = 0$ for all those F, then the distribution f is said to be zero on U, $f(\boldsymbol{x}) = 0$ on U. The **support of a distribution** f is the smallest closed set in $\mathbb{R}^n$ outside of which f is zero. Since for a bounded functional f on $\mathcal{E}$ the value (2.31) must be finite for all $F \in \mathcal{E}$, *$\mathcal{E}^*$ is the space of distributions with compact support*. (Dirac's δ-function and its derivatives have one-point support.)

Another most important case in physics regards **Fourier transforms of distributions**. Consider the subspace $\mathcal{S}$ of **rapidly decaying functions** of the class $C^\infty(\mathbb{R}^n, \mathbb{C})$ for which for every k and $\boldsymbol{m}$

$$\sup_{\boldsymbol{x}} |\boldsymbol{x}^{\boldsymbol{m}} D^k F(\boldsymbol{x})| < \infty, \quad \boldsymbol{x}^{\boldsymbol{m}} = \prod_{i=1}^{n} (x^i)^{m_i}, \quad D^k F \text{ like in (2.29).}$$

It is a topological vector space with the family of seminorms

$$p_{k,P}(F) = \sup_{\boldsymbol{x}} |P(\boldsymbol{x}) D^k F(\boldsymbol{x})|, \quad P : \text{polynomial in } \boldsymbol{x}. \tag{2.32}$$

Clearly, $\mathcal{S}$ is closed with respect to the operation with differential operators with polynomial coefficients. Since obviously $\mathcal{S} \subset C^\infty(\mathbb{R}^n, \mathbb{C}) \cap C_\infty(\mathbb{R}^n, \mathbb{C})$ (p. 33), $C_0^\infty(\mathbb{R}^n, \mathbb{C})$ is dense in $\mathcal{S}$ in the topology (2.32) of $\mathcal{S}$. In fact, *$\mathcal{S}$ is a complete (in the topology of $\mathcal{S}$) subspace of $\mathcal{E}(\mathbb{R}^n, \mathbb{C})$; it is again a Fréchet space*. The Fourier transform of a function of $\mathcal{S}$ is

$$\begin{aligned} (\mathcal{F}F)(\boldsymbol{k}) &= \frac{1}{(2\pi)^{n/2}} \int d^n x\, e^{-i(\boldsymbol{k}\cdot\boldsymbol{x})} F(\boldsymbol{x}), \\ F(\boldsymbol{x}) = (\bar{\mathcal{F}}(\mathcal{F}F))(\boldsymbol{x}) &= \frac{1}{(2\pi)^{n/2}} \int d^n x\, e^{i(\boldsymbol{x}\cdot\boldsymbol{k})} (\mathcal{F}F)(\boldsymbol{k}). \end{aligned} \tag{2.33}$$

Depending on context, the prefactor may be defined differently. It can be shown that $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is an isomorphism and $\mathcal{F}\bar{\mathcal{F}} = \mathrm{Id}_{\mathcal{S}}$, that is $\mathcal{F}^{-1} = \bar{\mathcal{F}}$.

The dual $\mathcal{S}^*$ of $\mathcal{S}$ is the space of **tempered distributions**, $\mathcal{S}^* \supset \mathcal{E}^*$. It is a module on the ring of polynomials (see Compendium), and is closed under differentiation. The Fourier transform in $\mathcal{S}^*$ is defined through the Fourier transform in $\mathcal{S}$ as

$$\langle \mathcal{F}f, F \rangle = \langle f, \mathcal{F}F \rangle. \tag{2.34}$$

Again, $\mathcal{F} : \mathcal{S}^* \to \mathcal{S}^*$ is an isomorphism, $\mathcal{F}\overline{\mathcal{F}} = \mathrm{Id}_{\mathcal{S}^*}$. If $f(\boldsymbol{k}) \equiv 1 \in \mathcal{S}^*$ is considered as a tempered distribution, then $\mathcal{F}f = (2\pi)^{n/2}\delta_0$.

A simple result relevant in the theory of Green's functions is the **Paley–Wiener theorem**: *The Fourier transform of a distribution with compact support on $\mathbb{R}^n$ can be extended into an analytic function on $\mathbb{C}^n$.*

Proofs of the above and more details can be found in textbooks of functional analysis, for instance [2]. (Closely related is also the theory of generalized solutions of partial differential equations, which are elements of Sobolev spaces.)

2.5 Connectedness, Homotopy

So far, the focus was mainly on the local topological structure which can be expressed in terms of neighborhood bases of points, although the concepts of vector space and of compactness and in particular of partition of unity provide a link to global topological properties. Connectedness has the focus on global properties, though with now and then local aspects. Intuitively, connectedness seems to be quite simple. In fact, it is quite touchy, and one has to distinguish several concepts.

A topological space is called **connected**, if it is not a union of two disjoint non-empty open sets; otherwise it is called **disconnected** (Fig. 2.6). Connectedness is equivalent to the condition that it is not a union of two disjoint non-empty closed sets, and also to the condition that the only open-closed sets are the empty set and the space itself. A subset of X is connected, if it is connected as the topological subspace with the relative topology; it need neither be open nor closed in the topology of X (cf. the definition of the relative topology). *If A is connected then every A' with $A \subset A' \subset \overline{A}$ is connected* (exercise).

Caution: Two disjoint sets which are not both open or both closed may have common boundary points being points of one of the sets and hence their union may be connected. The union of disjoint sets need not be disconnected.

The **connected component** of a point x of a topological space X is the largest connected set in X containing x. The relation $R(x, y)$: (y belongs to the connected

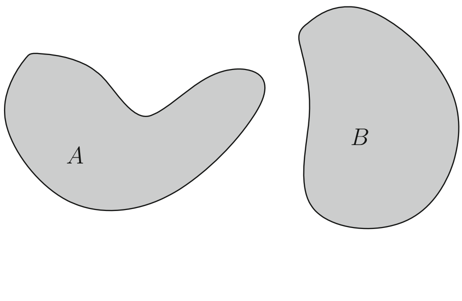

Fig. 2.6 Two connected sets A and B the union of which is disconnected

component of x) is an equivalence relation. The elements of the quotient space X/R are the connected components of X.

A topological space is called **totally disconnected**, if its connected components are all its one point sets $\{x\}$. Let $\pi : X \to X/R$ be the canonical projection onto the above quotient space X/R. The **quotient topology** of X/R is the finest topology in which π is continuous. Its open (closed) sets are the sets B for which $\pi^{-1}(B)$ is open (closed) in X. X/R is totally disconnected in the quotient topology.

Every set X is connected in its trivial topology and totally disconnected in its discrete topology. The rational line $\mathbb{Q}$ in the relative metric topology as a subset of $\mathbb{R}$ is totally disconnected. Indeed, let $\alpha < \beta$ be two rational numbers and let $\gamma, \alpha < \gamma < \beta$ be an irrational number. Then, $]-\infty, \gamma[$ and $]\gamma, +\infty[$ are two disjoint open intervals of $\mathbb{Q}$ the union of which is $\mathbb{Q}$. Hence, no two rational numbers belong to the same connected component of $\mathbb{Q}$. This example shows that the topology in which a space is totally disconnected need not be the discrete topology. In $\mathbb{Q}$, every one point set is closed (since $\mathbb{Q}$ as a metric space is Hausdorff) but not open. Open sets of $\mathbb{Q}$ are the rational parts of open sets of $\mathbb{R}$.

The image $F(A)$ of a connected set A in a continuous mapping is a connected set. Indeed, if $F(A)$ would consist of disjoint open sets then their preimages would be disjoint open sets constituting A. On the other hand, the preimage $F^{-1}(B)$ of a connected set B need not be connected (construct a counterexample). However, as connectedness is a topological property, a homeomorphism translates connected sets into connected sets in both directions. Check that, *if X is connected and Y is totally disconnected, for example if Y is provided with the discrete topology, then the only continuous functions $F : X \to Y$ are the constant functions on X*.

Let R be any equivalence relation in the topological space X. Since the canonical projection $\pi : X \to X/R$ is continuous in the quotient topology, it follows easily that *if the topological quotient space X/R is connected and every equivalence class in X with respect to R is connected, then X is connected.*

A topological space X is disconnected, iff there exists a continuous surjection onto a discrete two point space. (The target space may be $\{0, 1\}$ with the discrete topology; then, some of the connected components are mapped onto $\{0\}$ and some onto $\{1\}$.)

The topological product of non-empty spaces is connected, iff every factor is connected.

Proof Although the theorem holds for any number of factors, possibly uncountably many in Tichonoff's product, here only the case of finitely many factors is considered. (Though the proof works in the general case, only Tichonoff's product was not introduced in our context.) Let X_i be the factors of the product space X and $\pi_i : X \to X_i$ the canonical projections. Since these are continuous in the topological product, if X is connected, then every X_i as the image of X in a continuous mapping is connected. Now, assume that all X_i are connected but X is not. Then, there is a continuous surjection F of X onto $\{0, 1\}$. Let for some $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$, $\bar{x}_i \in X_i$, $F(\bar{x}) = 0$. Consider the subset $(x_1, \bar{x}_2, \ldots, \bar{x}_n)$, where x_1 runs through X_1, and the restriction of F on this subset. This restriction is a continuous function on X_1 and hence is $\equiv 0$ since X_1 is connected. Starting from every point of this subset, let now x_2 run through X_2 to obtain again $F \equiv 0$ for the restriction of F. After n steps, $F \equiv 0$ on X in contradiction to the assumption that F is surjective. $\square$

A concept seemingly related to connectedness but in fact independent is local connectedness. A topological space is called **locally connected**, if every point has a neighborhood base of connected neighborhoods. (Not just one neighborhood, *all* neighborhoods of the base must be connected.)

A connected space need not be locally connected. For instance, consider the subspace of $\mathbb{R}^2$ consisting of a horizontal axis and vertical lines through all rational points on the horizontal axis, in the relative topology deduced from the usual topology of the $\mathbb{R}^2$. It is connected, but no point off the horizontal axis has a neighborhood base of only connected sets. (Compare the above statement on $\mathbb{Q}$.) On the other hand, every discrete space with more than one point, although it is totally disconnected, is locally connected! Indeed, since every one point set is open and connected in this case, it forms a connected neighborhood base of the point. (Check it.) This seems all odd, nevertheless local connectedness is an important concept.

A topological space is locally connected, iff every connected component of an open set is an open set. This is not the case in the above example with the vertical lines through rational points of a horizontal axis, since the connected components of open sets off the horizontal axis are not open.

Proof of the statement Pick any point x and any neighborhood of it and consider the connected component of x in it. Since it is open, it is a neighborhood of x. Hence, x has a neighborhood base of connected sets, and the condition of the theorem is sufficient. Reversely, let A be an open set in a locally connected space, A' one of its connected components and x any point of A'. Let U be a neighborhood of x in A. It contains a connected neighborhood of x which thus is in A'. Hence, x is an inner point of A' and, since x was chosen arbitrarily, A' is open. $\square$

As a consequence, a locally connected space is a collection of its connected components *which are all open-closed.*

A topological quotient space of a locally connected space is locally connected.

Proof Let X be locally connected and let $\pi : X \to X/R$ be the canonical projection. Let $U \subset X/R$ be an open set and U' one of its connected components. Let $x \in \pi^{-1}(U')$, and let A be the connected component of x in $\pi^{-1}(U)$. Then, $\pi(A)$ is connected (since π is continuous) and contains $\pi(x)$. Hence, $\pi(A) \subset U'$ and $A \subset \pi^{-1}(U')$. Since X is locally connected and $\pi^{-1}(U)$ is open (again because π is continuous), $\pi^{-1}(U')$ is also open due to the previous theorem. Now, by the definition of the quotient topology, U' is also open, and the previous theorem in the opposite direction says that X/R is locally connected. $\square$

The subsequently discussed further concepts of connectedness are based on **homotopy**. Let $I = [0, 1]$ be the closed real unit interval. Two continuous functions

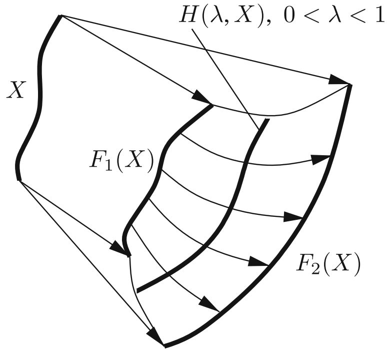

Fig. 2.7 Homotopic functions F_1 and F_2

F_1 and F_2 from the topological space X into the topological space Y are called **homotopic**, $F_1 \cong F_2$, if there exists a continuous function $H : I \times X \to Y$: $H(0, \cdot) = F_1$, $H(1, \cdot) = F_2$. H is called the homotopy translating F_1 into F_2 (Fig. 2.7). Since its definition is only based on the existence of continuous functions, homotopy is a purely topological concept.

The F_i may be considered as points in the functional space $C^0(X, Y)$. Then, $H(\lambda, \cdot)$, $0 \le \lambda \le 1$ is a path in $C^0(X, Y)$ from F_1 to F_2. If X and Y are normed vector spaces or manifolds, sometimes, in a narrower sense, the functions F_i, H are considered to be C^n-functions, $0 \le n \le \infty$. One then speaks of a C^n-homotopy. Of course, every C^n-homotopy is also a C^m-homotopy for $m \le n$. Homotopy is the C^0-homotopy. In the following statements homotopy may be replaced by C^n-homotopy with slight modifications in the construction of products H_2H_1 (see for instance [4, $VI.8]).

The product $H_2\ H_1$ of two homotopies, H_1 translating F_1 into F_2 and H_2 translating F_2 into F_3, may be introduced as a homotopy translating F_1 into F_3 in the following natural way by concatenating the two translations:

$$(H_2H_1)(\lambda, x) = \begin{cases} H_1(2\lambda, x) & \text{for } 0 \le \lambda \le 1/2 \\ H_2(2\lambda - 1, x) & \text{for } 1/2 \le \lambda \le 1. \end{cases}$$

Hence, if $F_1 \cong F_2$ and $F_2 \cong F_3$, then also $F_1 \cong F_3$. This means that homotopy is an equivalence relation among continuous functions. The corresponding equivalence classes $[F]$ of functions F are called **homotopy classes**. If a homeomorphism P of X onto itself is homotopic to the identity mapping $P \cong \mathrm{Id}_X$, then $F \circ P \cong F$ (exercise).

Two topological spaces X and Y are called **homotopy equivalent**, if there exist continuous functions $F : X \to Y$ and $G : Y \to X$ so that $G \circ F \cong \mathrm{Id}_X$ and $F \circ G \cong \mathrm{Id}_Y$. Two homeomorphic spaces are also homotopy equivalent, the inverse is, however, in general not true. A topological space is called **contractible**, if it is homotopy equivalent to a one point space. For instance, every topological vector space is contractible. The homotopy class of a constant function mapping X to a single point is called the **null-homotopy class**.

Of particular interest are the homotopy classes of functions from n-dimensional unit spheres S^n into topological spaces X possibly with a topological group structure. The latter means that the points of X form a group (with unit element $e \in X$) and the group operations are continuous. The unit sphere S^n may be considered as the set of points $\boldsymbol{s} \in \mathbb{R}^{n+1}$ with $\sum_{i=1}^{n+1}(s^i)^2 = 1$. S^0 is the two point set $S^0 = \{-1, 1\}, S^1$ is the circle, S^2 is the ordinary sphere, and so on. For $-1 < s^1 < 1$, the points $(s^2, \ldots, s^{n+1})$ with coordinates on $S^n, n > 0$, form an $(n-1)$-dimensional sphere (of radius r depending on s^1).

The case $n = 0$ is special and is treated separately. A topological space X is called **pathwise connected** (also called arcwise connected), if for every pair (x, x') of points of X there is a continuous function $H : I \to X, H(0) = x, H(1) = x'$. For a general topological space X, pathwise connectedness of pairs of points is an equivalence relation, and the equivalence classes are the **pathwise connected components** of X. If X is pathwise connected, then it is connected (exercise). The inverse is not in general true. Let X be the union of the sets of points $(x, y) \in \mathbb{R}^2$ with $y = \sin(1/x)$ and $(0, y)$, $y \in \mathbb{R}$ in the relative topology as a subset of $\mathbb{R}^2$. It is connected, but points with $x = 0$ and $x \neq 0$ are not pathwise connected. (Points $(0, y)$ with $|y| \leq 1$ are also not locally connected.) X is **locally pathwise connected**, if every point has a neighborhood base of pathwise connected sets. If X is locally pathwise connected, then it is locally connected, but again the inverse is not in general true.

For the following, $n \geq 1$, and *until otherwise stated, X is considered pathwise connected.* A homeomorphism between the sphere S^n, $n \geq 1$ and the n-dimensional unit cube with a particular topology is needed. Consider the open unit cube $I^n = \{\boldsymbol{x} | -1/2 < x^i < 1/2\}$ with its usual topology and its one point compactification $\overline{I^n}$, obtained by identifying the surface ∂I^n of I^n with the additional point $\boldsymbol{x}_\infty$ of $\overline{I^n}$. $\overline{I^n}$ is obviously homeomorphic to the one point compactification $\overline{\mathbb{R}^n}$ of $\mathbb{R}^n$, but it is also homeomorphic to S^n where a homeomorphism may be considered which maps $\boldsymbol{x}_\infty \in \overline{I^n}$ and $\boldsymbol{s}_0 = (1, 0, \ldots, 0) \in S^n$ onto each other. For $n = 1$ a homeomorphism between the unit circle and $\overline{\mathbb{R}}$ is obvious, for $n = 2$ it is a stereographic projection of the unit sphere S^2 onto the one-point compactified plane $\overline{\mathbb{R}^2}$. A similar mapping for $n > 2$ is easily found (exercise). The homeomorphism between S^n and $\overline{I^n}$ which maps $\boldsymbol{x}_\infty \in \overline{I^n}$ and $\boldsymbol{s}_0 = (1, 0, \ldots, 0) \in S^n$ onto each other is denoted by $\boldsymbol{P}$.

A word on notation her: x, x_0 denote points of X not having themselves coordinates since X in general is not a vector space; $\boldsymbol{x}$, $\boldsymbol{x}_\infty$ denote points in $\overline{I^n} \subset \mathbb{R}^n$ having coordinates $x^1, x^2, \ldots, x^n$ (not unique for $\boldsymbol{x}_\infty$); $\boldsymbol{s}, \boldsymbol{s}_0$ denote points on $S^n \subset \mathbb{R}^{n+1}$ having coordinates $s^1, s^2, \ldots, s^{n+1}$, $\sum_i (s^i)^2 = 1$.

Now, fix x_0 in the topological space X and consider the class $C_n(x_0)$ of continuous functions $F : S^n \to X$ with $F(\boldsymbol{s}_0) = x_0$ fixed. Denote the homotopy classes of functions $F \in C_n(x_0)$ by $[F]$. It is not the whole homotopy class of F in X, because for the group construction below it is necessary that the mapping of $\boldsymbol{s}_0 \mapsto x_0$ is fixed in every function F. The mapping F can be composed of two steps

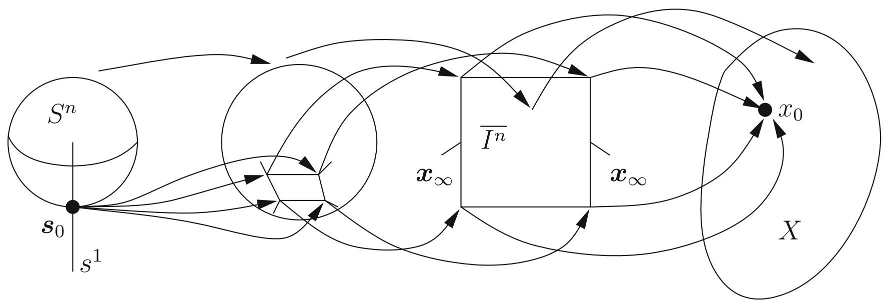

Fig. 2.8 Mapping of S^n onto $\overline{I^n}$ and $\overline{I^n}$ into X. It is visualized how the point s_0 is expanded into the square x_∞ which frames the image I^n of $S^n \setminus \{s_0\}$, and then x_∞ is mapped to x_0

(Fig. 2.8): first map S^n homeomorphically onto $\overline{I^n}$ by $\boldsymbol{P}$, implying $s_0 \mapsto \boldsymbol{x}_\infty$, and then map $\overline{I^n}$ into X by the continuous function $\tilde{F}$ with $\boldsymbol{x}_\infty \mapsto x_0$. Because $\boldsymbol{P}$ is a bijection, there is also a bijection between $\tilde{F}$ and $F = \tilde{F} \circ \boldsymbol{P}$, and $F(s_0) = x_0$.

This composition allows to explicitly define a group structure in the set of homotopy classes $[F]$ in the following way: For any two $C_n(x_0)$-functions F_1 and F_2 define a product $F_2F_1 \in C_n(x_0)$ by

$$(\tilde{F}_2\tilde{F}_1)(\boldsymbol{x}) = \begin{cases} \tilde{F}_1(2x^1 + 1/2, x^2, \ldots, x^n) & -1/2 \le x^1 \le 0 \\ \tilde{F}_2(2x^1 - 1/2, x^2, \ldots, x^n) & 0 \le x^1 \le 1/2, \end{cases} \quad (2.35)$$
$$F_2F_1 = (\tilde{F}_2\tilde{F}_1) \circ \boldsymbol{P}.$$

$(\tilde{F}_2\tilde{F}_1)$ is continuous, since the two functions $\tilde{F}_1$ and $\tilde{F}_2$ are glued together where $\tilde{F}_1(1/2, \ldots) = \tilde{F}_1(\boldsymbol{x}_\infty) = x_0 = \tilde{F}_2(\boldsymbol{x}_\infty) = \tilde{F}_2(-1/2, \ldots)$. Note that $\tilde{F}$ is supposed continuous with respect to the topology of $\overline{I^n}$ in which the surface ∂I^n is contracted into one point $\boldsymbol{x}_\infty$. Moreover, for $x^1 = -1/2$ or $x^1 = 1/2$, that is $\boldsymbol{x} = \boldsymbol{x}_\infty$, (2.35) yields $(\tilde{F}_2\tilde{F}_1)(\boldsymbol{x}_\infty) = x_0$, hence $F_2F_1 \in C_n(x_0)$. True, also $(\tilde{F}_2\tilde{F}_1)(0, \ldots) = x_0$ which for $|x_i| < 1/2, i = 2, \ldots, n$ is not demanded in the class $C_n(x_0)$. The construct (2.35) effectively pinches the section $x^1 = 0$ of $\overline{I^n}$ for $n > 1$ into one point. Via $\boldsymbol{P}$, this section corresponds to a meridian S^{n-1} of S^n containing the pole s_0. By moving from F_2F_1 to the homotopy class $[F_2F_1]$, this additional restriction (the pinch) is released.

In particular for $n = 1$, $\overline{I^1}$ is the line of length 1 with its endpoints identified (loop); hence it can again be considered as a circle. The mapping $\boldsymbol{P}$ which maps the pole s_0 to the connected endpoints of the second circle is trivial in this case. The point $\boldsymbol{x} = 0$ corresponds to the diametrically opposed point of the circle. In a product (2.35) of two mappings, this point is also mapped to x_0 making the product into a double loop (Fig. 2.9). The final correct product definition in the set of homotopy classes $[F]$ of functions with base point $F(s_0) = x_0$ is

$$[F_2][F_1] = [F_2F_1]. \quad (2.36)$$

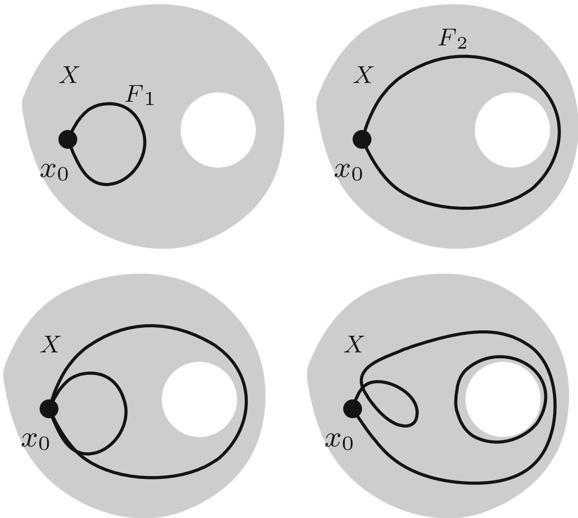

Fig. 2.9 Two loops $F_1, F_2 \in C_1(x_0)$ of the topological space X (shadowed area) and their product (2.35) (lower left panel). Also, another representative of $[F_2F_1]$ and a loop homotopic to F_2F_1 in X is shown (lower right panel). Since $[F_1] \cong E$ in this case, $[F_2F_1] \cong [F_1F_2] \cong [F_2]$

Next, having defined the product in (2.35, 2.36), it must be shown to be associative. Consider first

$$(\tilde{F}_3(\tilde{F}_2\tilde{F}_1))(\boldsymbol{x}) = \begin{cases} \tilde{F}_1(4x^1 + 3/2, \dots) & -1/2 \le x^1 \le -1/4 \\ \tilde{F}_2(4x^1 + 1/2, \dots) & -1/4 \le x^1 \le 0 \\ \tilde{F}_3(2x^1 - 1/2, \dots) & 0 \le x^1 \le 1/2 \end{cases}$$

and

$$((\tilde{F}_3\tilde{F}_2)\tilde{F}_1)(\boldsymbol{x}) = \begin{cases} \tilde{F}_1(2x^1 + 1/2, \dots) & -1/2 \le x^1 \le 0 \\ \tilde{F}_2(4x^1 - 1/2, \dots) & 0 \le x^1 \le 1/4 \\ \tilde{F}_3(4x^1 - 3/2, \dots) & 1/4 \le x^1 \le 1/2. \end{cases}$$

These two results differ only in a quite simple homeomorphism (piecewise linear in x^1, identity in the other coordinates) of $\overline{I^n}$ onto itself which is homotopic to $\mathrm{Id}_{\overline{I^n}}$. Hence, they are homotopic to each other (see p. 41). They also both map $\boldsymbol{x}_\infty$ to x_0. Thus, $[F_3]([F_2][F_1]) = [F_3(F_2F_1)] = [(F_3F_2)F_1] = ([F_3][F_2])[F_1]$.

If $\tilde{E}$ is the constant mapping $\tilde{E}(\boldsymbol{x}) \equiv x_0$ then obviously $e = [E]$ is a unity: $e[F] = [F] = [F]e$ for all $[F]$. Moreover, for $\tilde{F}_-(\boldsymbol{x}) = \tilde{F}(-x^1, x^2, \dots)$ (2.35) yields $[F_-][F] = e = [F][F_-]$. Indeed, $(\tilde{F}_-\tilde{F})(x^1, x^2, \dots) = (\tilde{F}_-\tilde{F})(-x^1, x^2, \dots)$: The image $(\tilde{F}_-\tilde{F})(\bar{I}^n)$ is a double layer in X. By symmetrically contracting the interval $-1/2 \le x^1 \le 1/2$ into $x^1 = 0$ with $x^2, \dots$ left constant $(\tilde{F}_-\tilde{F})(\bar{I}^n)$ shrinks continuously on itself into $x_0 = \tilde{E}(\bar{I}^n)$ by successive 'annihilation' of parts of the double layer. In total a group $\pi_n(X, x_0) = \{[F] | F \in C_n(x_0)\}$ is obtained with the group multiplication (2.36).

Now, consider *any* point x of the pathwise connected space X and a continuous path $H : I \to X$ with $H(0) = x_0$, $H(1) = x$. Given $F \in C_0(x_0)$, a function $F' \in C_0(x)$ may be constructed in the following manner:

$$\tilde{F}'(\boldsymbol{x}) = \begin{cases} \tilde{F}(2\boldsymbol{x}) & |x^i| \le 1/4,\ i = 1, \ldots, n, \\ H(t) & (2-t)\boldsymbol{x} \in \partial I^n,\ 0 \le t \le 1. \end{cases}$$

The base point x_0 of F is dragged along the path H to x. Apart from this path, the sets $\tilde{F}(\bar{I}^n)$ and $\tilde{F}'(\bar{I}^n)$ are the same which hence is also true for $F(\bar{I}^n)$ and $F'(\bar{I}^n)$. Moving F through its homotopy class $[F]$ with base point x_0 obviously also moves F' through its homotopy class $[F']$ with base point x. Moreover, it is easily seen (exercise) that $\tilde{F}_2\tilde{F}_1$ via H induces $\tilde{F}'_2\tilde{F}'_1$ for which $[F'_2F'_1] = [F'_2][F'_1]$. Hence, the mapping $\tilde{H} : [F] \mapsto [F']$ is a homomorphism of groups. Two concatenated paths H_1 and H_2 obviously induce a composition of homomorphisms $\tilde{H}_2 \circ \tilde{H}_1$. Concatenate now the path H with its reversed $H_-(t) = H(1-t)$. Then H_-H provides the identity map $\mathrm{Id}_{C_0(x_0)}$ while HH_- provides $\mathrm{Id}_{C_0(x)}$. $\tilde{H}$ and $\tilde{H}_-$ are thus inverse to each other, and the homomorphism $\tilde{H}$ is in fact an isomorphism. The groups $\pi_n(X, x)$ and $\pi_n(X, x_0)$ are isomorphic, or, in other words, $\pi_n(X, x_0) \approx \pi_n(X)$ does not depend on x_0. The group $\pi_n(X)$ is called the ***n*th homotopy group** of the pathwise connected topological space X.

Since the case $n = 1$ is of particular interest in the theory of integration on manifolds (see Chap. 5), $\pi_1(X)$ is called the **fundamental group** of X.

Formally, a '0-dimensional open cube' can be considered as a one point set $I^0 = \{\boldsymbol{x}\}$, and its one point 'compactification' (I^0 is of course also compact) as the discrete two point set $\overline{I^0} = \{\boldsymbol{x}, \boldsymbol{x}_\infty\}$. The homeomorphism $\boldsymbol{P}$ between $S^0 = \{-1, 1\}$ and $\overline{I^0}$ maps -1 to $\boldsymbol{x}$ and 1 to $\boldsymbol{x}_\infty$. Now, $F : S^0 \to X$ is a two point mapping, and $F \in C_0(x_0)$ means that $F(-1) = x$ where x is *any* point of X, and $F(1) = x_0$. The classes $[F]$ thus map -1 into the pathwise connected components of X, and x_0 does not play any role. *For a pathwise connected topological space X, $\pi_0(X) = \{e\}$ is trivial.*

By inspection of (2.35) it is seen that interchanging the factors in the multiplication amounts to interchanging the halves $x^1 \le 0$ and $x^1 \ge 0$ in $\overline{I^n}$. For $n > 1$, the positioning of these two halves relative to each other does not play a role because of the pinch of the section $x^1 = 0$ involved in (2.35). Therefore, the interchanging of the two halves can be provided by a homeomorphism of I^n onto itself which is also homotopic to the identity mapping: note that $\overline{I^n}$ is homeomorphic to a cylinder with axis perpendicular to the x^1-axis. Rotate it by 180° to transform continuously from the identity to the interchanging of the above two halves. *The groups $\pi_n(X)$, $n \ge 2$ of a pathwise connected topological space X are commutative*. For that reason, in the literature the group operation of homotopy groups is often denoted as addition instead of multiplication.

In the case $n = 1$ the interchanging may still be provided by a homeomorphism, however, the argument of deformation into a cylinder does not work any more, and the interchanging is not any more homotopic to the identity mapping. *The fundamental group $\pi_1(x)$ need not be commutative*. Consider for instance a two-dimensional space X with two holes and a loop first orbiting clockwise around the first hole and then counterclockwise around the second. Check that this loop is not homotopic to the loop with the sequence of orbiting interchanged.

If X itself has a group structure, that is, X *is a topological group* with multiplication denoted by a dot (to distinguish it from the multiplication (2.36)), and $x_0 = e$, then another product of $C_n(e)$-functions and the inverse of a $C_n(e)$-function may alternatively be defined by pointwise application of the group operations. The $C_n(e)$-unity is the constant mapping on e. Let $F_1 \cong F'_1$ and $F_2 \cong F'_2$ and consider the homotopies H_i translating F_i into F'_i, $(H_i(0,\cdot) = F_i,\, H_i(1,\cdot) = F'_i)$. Then $H_1 \cdot H_2$ is a homotopy translating $F_1 \cdot F_2$ into $F'_1 \cdot F'_2$, hence $[F_1 \cdot F_2] = [F'_1 \cdot F'_2]$: the group multiplication in X is compatible with the homotopy class structure of $C_n(e)$ and the multiplication $[F_1] \cdot [F_2]$ is properly defined. Clearly, $e = [E]$ is also the unity for the dot multiplication. Moreover, with (2.35), $\tilde{F}_1\tilde{F}_2 = (\tilde{F}_1\tilde{E}) \cdot (\tilde{E}\tilde{F}_2)$ is easily verified (check it). The conclusion is $[F_1][F_2] = [F_1] \cdot [F_2]$: the dot-multiplication yields again the same homotopy group $\pi_n(X,e)$ *of the pathwise connected component of* e *in* X as previously. Since the multiplication (from left or right) with any element x of the component X^e of e in X yields a translation of that component which is also a homeomorphism of that component X^e onto itself, $\pi_n(X,e) \approx \pi_n(X,x) \approx \pi_n(X^e)$ for any x of the component of e in X.

However, if the topological group X *is not pathwise connected*, in a wider sense the homotopy group $\pi_n(X)$ with the dot-multiplication can still be constructed. In this case, $\pi_0(X)$ is non-trivial, and the elements of $\pi_0(X)$ are in a one–one correspondence with the pathwise connected components of X. Let $x \notin X^e$ be a group element not in the pathwise connected component of e, and let $x_0 \in X^e$, that is, there is a continuous path connecting x_0 with e. Since in a topological group the group operations are continuous, it follows that there is a continuous path from $x \cdot x_0$ to $x \cdot e = x$; $x \cdot x_0 \in X^x$, and likewise $x_0 \cdot x \in X^x$. It is easily seen that all pathwise connected components of a group X are homeomorphic to each other (exercise). It follows further that there is a continuous path connecting $x \cdot x_0 \cdot x^{-1}$ with $x \cdot e \cdot x^{-1} = e$. Hence, $x \cdot x_0 \cdot x^{-1} \in X^e$ for every $x \in X$ and every $x_0 \in X^e : X^e$ is an invariant subgroup of X. It is easily seen that $X/X^e \approx \pi_0(X)$. On the other hand, $x \cdot x_1 \cdot x^{-1} \mapsto x'_1, x_1 \in X, x \in X$ is an automorphism of X for any fixed x which, as was seen, transforms pathwise connected components of X into themselves.

Consider $C_n(x)$-functions F from the S^n-sphere into X with *any* base point x, not necessarily in X^e. The homotopy classes $[F]$ in $C_n(x)$ form a larger group $\pi_n(X)$ which now is only defined with the group multiplication $[F_1] \cdot [F_2]$. The above considered automorphism of X yields in a canonical way an automorphism of $\pi_n(X)$. Denote the elements of $\pi_0(X)$ by $[H]$; then the anticipated automorphism is given by

$$[F]' = [H] \cdot [F] \cdot [H]^{-1}, \quad [F] \in \pi_n(X),\ [H] \in \pi_0(X). \tag{2.37}$$

If $[F] \subset C_n(e)$ then $[F]' \subset C_n(e)$, hence, $\pi_n(X^e)$ is an invariant subgroup of $\pi_n(X)$, and $\pi_0(X) \approx \pi_n(X)/\pi_n(X^e)$. Because of the above discussed structure of the pathwise connected classes of X, obviously also $\pi_n(X^e) = \pi_n(X)/\pi_0(X)$ and hence

$$\pi_n(X) = \pi_0(X) \times \pi_n(X^e), \quad n > 0 \tag{2.38}$$

for the **homotopy groups of a topological group** X. They can be quite different from $\pi_n(X^e)$ (and need not be commutative for any $n \geq 0$ since $\pi_0(X)$ need not be commutative any more).

A topological space X is called n-**connected** (sometimes called n-simple), if every continuous image in X of the n-dimensional sphere S^n is contractible. A topological group X is n-connected, if $\pi_n(X) \approx \pi_0(X)$. An n-connected space need not be connected. A 0-connected space is pathwise connected, a 1-connected space is called **simply connected**.[2] Although n-connectedness is very similarly defined for different n, these properties are largely unrelated (except for the role of π_0). Some authors apply n-connectedness only to pathwise connected spaces X. However, for many applications this is an unnecessary restriction.

Some examples are given without proof. Some of them are intuitively clear. (1) A convex open subspace of a topological vector space is n-connected for *any* $n \geq 0$. (2) The sphere S^n or the complement to the origin in $\mathbb{R}^{n+1}$ is k-connected for $0 \leq k \leq n-1$; for $n > 1$ it is simply connected. (3) $\pi_n(S^n) = \mathbb{Z}$ (as an additively written Abelian group). For an integer $m \in \mathbb{Z} = \pi_n(S^n), |m|$ is the cardinality of $F^{-1}(x)$ for any $x \in S^n$. It is called the **degree of** the **mapping** F. (4) $\pi_n(S^m)$, $n > m$ is a largely unsolved problem although many special cases have meanwhile been compiled; $\pi_3(S^2) = \mathbb{Z}$ is a theorem by Hopf, and $\pi_2(S^1) = 0$ is easily understood. (5) For the torus $\mathbb{T}^2$ (see Fig. 1.3), $\pi_1(\mathbb{T}^2) = \mathbb{Z} \times \mathbb{Z}$. One integer of $(m_1, m_2) \in \mathbb{Z} \times \mathbb{Z}$ counts the oriented windings around the circumference of the tire, and the other those around its cross section.

These concepts are further exploited in Chaps. 5 and 8. Although the physical relevance of homotopy was anticipated already by Poincaré, it turned out to be one of the most difficult and unsolved tasks of topology to calculate the homotopy groups of certain manifolds and to exploit them for classification. It was already known to Poincaré that every compact simply connected two-dimensional manifold without boundary is homeomorphic to the sphere S^2. His conjecture that the same is true in three dimensions and every compact simply connected three-dimensional manifold without boundary is homeomorphic to the 3-sphere S^3 withstood hard attempts by able mathematicians for hundred years to prove it and was eventually proved only quite recently by G. Perelman.

---

[2] There is a more general definition of simple connectedness and fundamental group in terms of covering space. For pathwise connected locally pathwise connected spaces X it is equivalent to the definition given here [6].

2.6 Topological Charges in Physics

In quantum physics, thermodynamic phases are characterized by order parameters: the particle densities of various particles, atom displacements of crystalline solids, the magnetization density vector, the anomalous Green function of the superconducting or superfluid state and so on. In an inhomogeneous, in particular defective state those order parameters are functions of space (and maybe time). The various defects can often be classified by discrete **topological charges**, and then those classes turn out to be stable: because of the discrete nature of the charges there is no continuous transformation of one class into another. The topological charges are often generating elements of homotopy groups.

Consider as a simple example a superconducting state in three dimensions penetrated by a vortex line. The space X of the superconducting state is $\mathbb{R}^3$ with the vortex line cut out. It is homotopy equivalent to a circle S^1 around the vortex line. The order parameter $\Delta = |\Delta|e^{i2\pi\phi}$ of a conventional superconducting state (spin singlet s wave) is a complex number having a phase ϕ the gradient of which is proportional to the supercurrent while the absolute value $|\Delta|$ is the gap which is fixed for a given material and for given temperature and pressure. A constant phase factor is irrelevant, the state is degenerate with respect to an arbitrary complex phase factor. The loop S^1 in the complex plane of all phase factors is the order parameter space Γ of degenerate states in that case. With a defect present in X, the order parameter in general will be position dependent with values out of Γ. This position dependence defines a mapping $F : X \to \Gamma$. Since Δ is a well defined function on X, the gradient $\partial\phi/\partial\boldsymbol{x}$ of the phase must integrate along any closed loop to an integer, $\oint \boldsymbol{ds} \cdot (\partial\phi/\partial\boldsymbol{x}) = \text{integer}$, and this integer must be the same for all homotopy equivalent loops. On a loop not encircling the vortex line this integer must be zero, since the loop may be continuously contracted within X to a point, and a non-zero integer cannot continuously be changed to zero. On a loop once encircling the vortex line the integral of the gradient of the phase ϕ may be any integer N characterizing the vortex line. For a loop m times winding around the vortex line it then is Nm. N is the number of magnetic flux quanta in the vortex line. It generates a group of elements Nm with $m \in \mathbb{Z}$. This group is obviously isomorphic with the group $\mathbb{Z}$, which in this case is the fundamental group $\pi_1(\Gamma = S^1)$ of homotopy classes of mappings from S^1 which is homotopy equivalent to X into $\Gamma = S^1$.

On a discrete lattice, the sum of unit lattice periods along a loop is similar to a phase and must be an integer number of lattice vectors along the loop. For a loop enclosing a defect free region of the crystal this sum is zero. For a loop around a displacement line this is the Burgers vector of the displacement. Here the space X of the crystalline phase is again the same as above and is again homotopy equivalent to the circle S^1, this time around the displacement line. Any loop yields m times the Burgers vector.

Such situations will in more generality and more detail be considered in Chap. 8. Here, some principal remarks are in due place. The Hamiltonian of a macroscopic system has in general a number of symmetries, it is invariant with respect to transformations of a symmetry group G, translational, rotational invariance, gauge symmetries and others. Some of the symmetries may be approximate, but obeyed to a sufficient level of accuracy. For instance in a rare gas liquid the coupling of the nuclear spin with the rotational motion is so weak that invariance with respect to spatial rotation and spin rotation may be considered separately. At sufficiently high temperature, the state of the system is completely disordered, so that its thermodynamic (macroscopic) variables are invariant under the symmetry transformations of G. The thermodynamic state γ fulfills the relation $\gamma = g\gamma$ for all $g \in G$ and is thus uniquely determined. In the course of lowering the temperature, phase transitions may take place with developing non-zero order parameters so that now γ is not any more invariant with respect to all symmetry transformations g of G, but may still be invariant with respect to a subgroup H of G. Then, γ generates an orbit $\{g\gamma | g \in G\}$ which is isomorphic to the quotient space $\Gamma = G/H$ of left cosets of H in G. It is this quotient space which figures as the order parameter space Γ in the above considerations.

In the above example of a line defect in $\mathbb{R}^3$ it was essential only that the defect free space X was homotopically equivalent to a circle S^1. The number of topological 'charges' of the defect is then equal to the number of generators of the homotopy group $\pi_1(\Gamma)$ (one in the above cases). The same would be true for a point defect in $\mathbb{R}^2$ or a line defect propagating in time (defect world sheet) in four-dimensional space–time. For a point defect in $\mathbb{R}^3$, X is homotopy equivalent to a sphere S^2 enclosing the defect, and hence the number of its topological charges is equal to the number of generators of $\pi_2(\Gamma)$.

In general, the number of topological charges of a defect of codimension d in a state with order parameter space Γ present in an n-dimensional position space (i.e., the dimension of the defect is $n - d$) is equal to the number of generators of the homotopy group $\pi_{d-1}(\Gamma)$.

In order to develop a non-zero topological quantum number (non-trivial topological charge), a defect of codimension d in a state with order parameter space Γ must have a non-trivial homotopy group $\pi_{d-1}(\Gamma)$. Consider as an example an isotropic magnetically polarizable material. The Hamiltonian does not prefer any direction in space, besides translational invariance which need not be considered here (it assures that a magnetization vector smoothly depending on position has low energy) the continuous symmetry group is $SO(3)$ (cf. Chap. 6). At sufficiently high temperature, above the magnetic order temperature, the magnetic polarization is disordered on an atomic scale and the state γ is invariant: $\gamma = g\gamma$ for all $g \in G = SO(3)$. Below the ordering temperature the magnetization density vector is non-zero. Its absolute value is determined by the material, temperature and pressure. Its direction may be arbitrary, and all directions are energetically degenerate. Smooth long wavelength changes of direction have low excitation energy. If the non-zero magnetization points in a certain direction, the state is still invariant with respect to rotations of the group $H = SO(2)$ around the axis of polarization. The order parameter space is

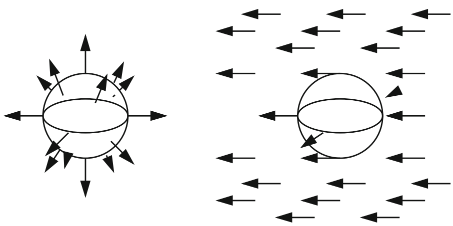

Fig. 2.10 Point defect of (a) an isotropic magnetic material, so-called hedgehog, and (b) of an easy plane anisotropic magnetic material with no non-trivial topological charge possible

$SO(3)/SO(2)$ and consists of all vectors of a given length pointing in all possible spatial directions. Topologically this group is homeomorphic to the sphere S^2. Hence, $\Gamma = S^2$. For a point defect in 3-space (codimension 3), $\pi_2(S^2) = \mathbb{Z}$ (see end of last section). Hence, the point defect may have a non-trivial topological charge in this case.

A point defect is a small spot where the magnetization density vanishes. Outside of a sphere of a small radius it is again fully developed, but may for instance everywhere point in radial direction (Fig. 2.10a). The change of direction outside of this sphere is everywhere smooth, but there is no smooth transition into a homogeneously magnetized state with constant magnetization direction. This 'hedgehog' point defect has non-trivial topological charge and is stable: the defect cannot be resolved by smooth magnetization changes.

Consider now an anisotropic magnetic material of the type easy plane. Again the magnitude of the magnetization density vector is fixed at given temperature and pressure, but can only point in the directions within a plane, $\Gamma = SO(2) \sim S^1$. Now $\pi_2(S^1) = 0$: the sphere S^2 is simply connected and cannot be continuously wound around a circle. Hence no non-trivial topological charge of a point defect is possible in this case. From Fig. 2.10b it is easily inferred that no hedgehog-like structure is possible without singularity lines outside a sphere around the defect of the magnetization vector field of constant magnitude. From the singularity lines the magnetization density vector would point into all planar radial directions. If this is a linear defect, it is governed by $\pi_1(S^1) = \mathbb{Z}$, and a topological charge can exist on the linear defect in an easy plane magnet.

A point defect of codimension 4 in four-dimensional space–time would be capable of carrying a topological charge, if $\pi_3(\Gamma)$ is non-trivial. Just to mention it, the Belavin–Polyakov instanton of a Yang–Mills field is such a case even without a defect (Chap. 8).

Structures with topological charges may intrinsically exist without a material defect. Consider the plane $\mathbb{R}^2$ with a non-zero magnetization density which approaches a homogeneous magnetization density vector of a fixed direction at infinite distance from the origin of $\mathbb{R}^2$. This state may be considered as a state in the compactified plane $\overline{\mathbb{R}^2} \sim S^2$ which is homeomorphic to a sphere via the stereographic projection. Since the order parameter space Γ of an $SO(3)$ spin is

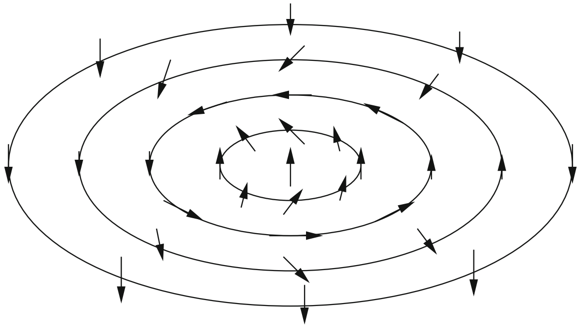

Fig. 2.11 Baby skyrmion on a planar magnet with magnetization density vector up in the center and down at infinity by a spiral rotation around the radial direction

also S^2, one has $\pi_2(\Gamma) = \pi_2(S^2) = \mathbb{Z}$, and hence there exists a topological charge. The corresponding magnetic state is called a 'baby **skyrmion**' and is the only skyrmion structure for which a picture can be drawn. It is shown in Fig. 2.11.

This state has a three-dimensional analogue since $\overline{\mathbb{R}^3} \sim S^3$ and $\pi_3(S^2) = \mathbb{Z}$ is the famous Hopf theorem. The corresponding Hopf mapping of S^3 onto S^2 is however not easy to draw. In general, skyrmions are special solitons in n dimensions corresponding to non-trivial homotopy groups $\pi_n(\Gamma)$. Originally, T. H. R. Skyrme proposed a subgroup of the product of the left and right chiral copies of $SU(N)$ as the order parameter space Γ to obtain local field structures as candidates of baryons in Yang–Mills field theories. For a more detailed discussion of the Hopf mapping and citations for further reading see [7].

More examples of topological charges can be found in [8].

The section is closed by a consideration of the topological stability of the **Fermi surface** of a Fermi liquid. (A more detailed discussion of Fermi surfaces is given in Sect. 5.9.) Again, first the two-dimensional case is considered which can easily be visualized. For a non-interacting isotropic Fermi gas, the single-particle Green function at imaginary frequency $\omega = ip_0$ is

$$G(ip_0, \boldsymbol{p}) = \frac{1}{ip_0 - v_F(p - p_F)}, \tag{2.39}$$

where $\boldsymbol{p}$ is the momentum vector, $p = |\boldsymbol{p}|$, p_F is the Fermi momentum, and v_F is the Fermi velocity. The energy dispersion close to the Fermi surface $p = p_F$ is $\varepsilon = v_F(p - p_F)$. States with $p < p_F$ have negative energies (measured from the chemical potential $\varepsilon = \varepsilon_F$) and are occupied, while states with $p > p_F$ have positive energies and are unoccupied. The Fermi surface $p = p_F$ in two-dimensional momentum space is a circle (Fig. 2.12) separating the occupied momentum region from the unoccupied one.

The Green function $G(ip_0, \boldsymbol{p})$ has a singularity line $p_0 = 0, p = p_F$ forming the Fermi surface and is otherwise a complex analytic function for imaginary frequencies. If one maps the contour C in the $(p_0, \boldsymbol{p})$-space onto the complex plane of G^{-1} with Re $G^{-1} = -v_F(p - p_F)$, Im $G^{-1} = p_0$, it maps the circle C onto the

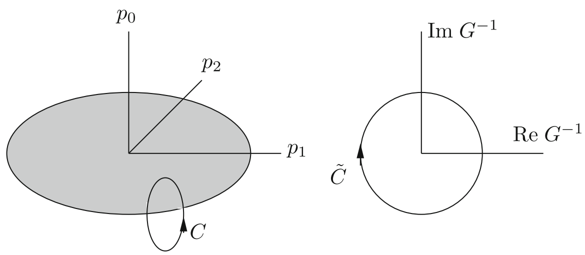

Fig. 2.12 *Left*: Fermi surface in two-dimensional momentum space and imaginary frequency axis with a loop C around the Fermi surface. *Right*: the corresponding loop $\tilde{C}$ of the complex function $G^{-1}(ip_0,\boldsymbol{p})$

circle $\tilde{C}$. Writing $G^{-1} = |G|^{-1}e^{-i\phi}$ it is seen that the phase of G increases by 2π when running around C, while for any loop not encircling the Fermi surface it returns to the start value. (This is like the phase of the superconducting order parameter when running around a vertex line.) The Fermi surface is like a defect line in momentum space.

If now the interaction between the particles is continuously switched on, the Green function changes smoothly. It cannot smoothly get rid of its denominator because of this topological charge on the Fermi surface, hence it must have the form

$$G(ip_0,\boldsymbol{p}) = \frac{Z}{ip_0 - v'_\mathrm{F}(p - p_\mathrm{F})}, \tag{2.40}$$

where Z is the spectral amplitude renormalization factor, and the Fermi velocity may change. (That p_F does not change is an independent result, the Luttinger theorem.) Hence the Fermi surface is topologically stabilized and can only disappear when Z becomes zero (which is only possible in a non-analytic way).

The only change for the case of three spatial dimensions is that now $\boldsymbol{p}$ is a 3-vector in the three-dimensional hyperplane of the four-dimensional frequency-momentum space of points $(p_0,\boldsymbol{p})$ for $p_0 = 0$, which contains the only singularities of (2.40) on the Fermi surface being now a 2-sphere. For every planar section in the three-dimensional momentum space through its origin, Fig. 2.12 visualizes further on the situation, and the Fermi surface is topologically stable.

A more general situation is present for electrons as spin 1/2 fermions in a crystalline solid instead of 'spinless fermions' in an isotropic medium which was considered so far. Here, the Green function is a complex valued matrix quantity indexed with band and spin indices. The change of its phase, normalized to 2π, as a complex number when going around a loop (contour integral of the gradient of the phase as considered in the case of a superconductor with a vertex line) is now to be replaced by the quantity

$$N = \mathrm{tr}\oint \frac{ds_p}{2\boldsymbol{i}} \cdot G(ip_0,\boldsymbol{p})\frac{\partial}{\partial p}G^{-1}(ip_0,\boldsymbol{p})$$

where the trace of the matrix product is to be formed, the contour integral is along the previous contour C, and $\partial/\partial p$ is the four-dimensional gradient in the frequency-momentum space. The dot means the scalar product of the line element vector with this gradient. This is the general structure of a homotopy invariant.[3]

Now, several sheets of Fermi surface may coexist of arbitrary shape. The shape may change when the interaction is tuned up and individual sheets may appear or disappear on the cost of other sheets. (If a Fermi radius shrinks to zero, in most cases the Fermi velocity also approaches zero, and the singularity disappears. Exceptions are so-called Dirac quasi-particles where the Fermi velocity remains non-zero in the Fermi points.) Nevertheless, between such changes the Fermi surface is topologically stable, and the only additional reason for its change is the vanishing of the spectral amplitude renormalization function $Z(p_0, \boldsymbol{p})$ on some part of the Fermi surface.

A much deeper analysis can be found in [9].

References

1. Reed M., Simon B.: Methods of Modern Mathematical Physics, Vol I. Functional Analysis Academic Press, New York (1973)
2. Yosida K.: Functional Analysis. Springer-Verlag, Berlin (1965)
3. Kolmogorov A., Fomin S.: Introductory Real Analysis. Prentice Hall, Englewood Cliffs NJ (1970)
4. Schwartz L.: Analyse Mathématique. Hermann, Paris (1967)
5. Eschrig H.: The Fundamentals of Density Functional Theory 1. Edition am Gutenbergplatz, Leipzig, p 226 (2003)
6. Choquet-Bruhat Y., de Witt-Morette C., Dillard-Bleick M.: Analysis, Manifolds and Physics, vol I: Basics. Elsevier, Amsterdam (1982)
7. Protogenov AP.: Knots and links in order parameter distributions of strongly correlated systems. Physics–Uspekhi **49**, 667–691 (2006)
8. Thouless DJ.: Toplogical Quantum Numbers in Nonrelativistic Physics. World Scientific, Singapore (1998)
9. Volovik GE.: The Universe in a Helium Droplet. Clarendon Press, Oxford (2003)

[3] The general expression for a topological charge enclosed by an n-sphere S^n (generator of the homotopy group $\pi_n(\Gamma)$) is $N_n = (n!|S^n|)^{-1} \int_{S^n} d\phi \wedge \cdots \wedge d\phi$, where $|S^n| = 2\pi^{(n+1)/2}/\Gamma((n+1)/2)$ is the volume of the n-sphere, $d\phi = \sum_{i=1}^{n} dx^i \partial\phi/\partial x^i$ is the 1-form of the gradient of the phase ϕ, and the $\wedge$-product has n factors, see later in Sect. 5.1.

http://www.springer.com/978-3-642-14699-2

Topology and Geometry for Physics
Eschrig, H.
2011, XII, 390 p. 60 illus., Softcover
ISBN: 978-3-642-14699-2