Excerpts from

Introduction to Modern Topology and Geometry

Anatole Katok

Alexey Sossinsky

Contents

CHAPTER 1

BASIC TOPOLOGY

Topology, sometimes referred to as "the mathematics of continuity", or "rubber sheet geometry", or "the theory of abstract topological spaces", is all of these, but, above all, it is a *language*, used by mathematicians in practically all branches of our science. In this chapter, we will learn the basic words and expressions of this language as well as its "grammar", i.e. the most general notions, methods and basic results of topology. We will also start building the "library" of examples, both "nice and natural" such as manifolds or the Cantor set, other more complicated and even pathological. Those examples often possess other structures in addition to topology and this provides the key link between topology and other branches of geometry. They will serve as illustrations and the testing ground for the notions and methods developed in later chapters.

1.1. Topological spaces

The notion of topological space is defined by means of rather simple and abstract axioms. It is very useful as an "umbrella" concept which allows to use the geometric language and the geometric way of thinking in a broad variety of vastly different situations. Because of the simplicity and elasticity of this notion, very little can be said about topological spaces in full generality. And so, as we go along, we will impose additional restrictions on topological spaces, which will enable us to obtain meaningful but still quite general assertions, useful in many different situations in the most varied parts of mathematics.

1.1.1. Basic definitions and first examples.

DEFINITION 1.1.1. A *topological space* is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a family of subsets of X (called the *topology* of X) whose elements are called *open sets* such that

(1) $\varnothing, X \in \mathcal{T}$ (the empty set and X itself are open),
(2) if $\{O_\alpha\}_{\alpha \in A} \subset \mathcal{T}$ then $\bigcup_{\alpha \in A} O_\alpha \in \mathcal{T}$ for any set A (the union of any number of open sets is open),
(3) if $\{O_i\}_{i=1}^k \subset \mathcal{T}$, then $\bigcap_{i=1}^k O_i \in \mathcal{T}$ (the intersection of a finite number of open sets is open).

If $x \in X$, then an open set containing x is said to be an (*open*) *neighborhood* of x.

We will usually omit $\mathcal{T}$ in the notation and will simply speak about a "topological space X" assuming that the topology has been described.

The complements to the open sets $O \in \mathcal{T}$ are called *closed* sets .

EXAMPLE 1.1.2. Euclidean space $\mathbb{R}^n$ acquires the structure of a topological space if its open sets are defined as in the calculus or elementary real analysis course (i.e a set $A \subset \mathbb{R}^n$ is open if for every point $x \in A$ a certain ball centered in x is contained in A).

EXAMPLE 1.1.3. If all subsets of the integers $\mathbb{Z}$ are declared open, then $\mathbb{Z}$ is a topological space in the so–called discrete topology.

EXAMPLE 1.1.4. If in the set of real numbers $\mathbb{R}$ we declare open (besides the empty set and $\mathbb{R}$) all the half-lines $\{x \in \mathbb{R} | x \geq a\}$, $a \in \mathbb{R}$, then we do not obtain a topological space: the first and third axiom of topological spaces hold, but the second one does not (e.g. for the collection of all half lines with positive endpoints).

EXAMPLE 1.1.5. Example 1.1.2 can be extended to provide the broad class of topological spaces which covers most of the natural situations.

Namely, a *distance function* or a *metric* is a function of two variables on a set X (i,e, a function of the Cartesian product $X \times X$ of X with itself) which is nonnegative, symmetric, strictly positive outside the diagonal, and satisfies the triangle inequality (see Definition 2.1.1). Then one defines an (open) ball or radius $r > 0$ around a point $x \in X$ as the set of all points at a distance less that r from X, and an open subset of X as a set which together with any of its points contains some ball around that point. It follows easily from the properties of the distance function that this defines a topology which is usually called the *metric topology*. Naturally, different metrics may define the same topology. We postpone detailed discussion of these notions till Chapter 2 but will occasionally notice how natural metrics appear in various examples considered in the present chapter.

The *closure* $\bar{A}$ of a set $A \subset X$ is the smallest closed set containing A, that is, $\bar{A} := \bigcap\{C \mid A \subset C \text{ and } C \text{ closed}\}$. A set $A \subset X$ is called *dense* (or *everywhere dense*) if $\bar{A} = X$. A set $A \subset X$ is called *nowhere dense* if $X \setminus \bar{A}$ is everywhere dense.

A point x is said to be an *accumulation point* (or sometimes *limit point*) of $A \subset X$ if every neighborhood of x contains infinitely many points of A.

A point $x \in A$ is called an *interior point* of A if A contains an open neighborhood of x. The set of interior points of A is called the *interior* of A and is denoted by $\operatorname{Int} A$. Thus a set is open if and only if all of its points are interior points or, equivalently $A = \operatorname{Int} A$.

A point x is called a *boundary point* of A if it is neither an interior point of A nor an interior point of $X \setminus A$. The set of boundary points is called the *boundary* of A and is denoted by ∂A. Obviously $\bar{A} = A \cup \partial A$. Thus a set is closed if and only if it contains its boundary.

EXERCISE 1.1.1. Prove that for any set A in a topological space we have $\partial \overline{A} \subset \partial A$ and $\partial(\operatorname{Int} A) \subset \partial A$. Give an example when all these three sets are different.

A sequence $\{x_i\}_{i\in\mathbb{N}} \subset X$ is said to *converge* to $x \in X$ if for every open set O containing x there exists an $N \in \mathbb{N}$ such that $\{x_i\}_{i>N} \subset O$. Any such point x is called a *limit* of the sequence.

EXAMPLE 1.1.6. In the case of Euclidean space $\mathbb{R}^n$ with the standard topology, the above definitions (of neighborhood, closure, interior, convergence, accumulation point) coincide with the ones familiar from the calculus or elementary real analysis course.

EXAMPLE 1.1.7. For the real line $\mathbb{R}$ with the discrete topology (all sets are open), the above definitions have the following weird consequences: any set has neither accumulation nor boundary points, its closure (as well as its interior) is the set itself, the sequence $\{1/n\}$ does not converge to 0.

Let $(X, \mathcal{T})$ be a topological space. A set $D \subset X$ is called *dense* or *everywhere dense* in X if $\bar{D} = X$. A set $A \subset X$ is called *nowhere dense* if $X \setminus \bar{A}$ is everywhere dense.

The space X is said to be *separable* if it has a finite or countable dense subset. A point $x \in X$ is called *isolated* if the one–point set $\{x\}$ is open.

EXAMPLE 1.1.8. The real line $\mathbb{R}$ in the discrete topology is *not* separable (its only dense subset is $\mathbb{R}$ itself) and each of its points is isolated (i.e. is not an accumulation point), but $\mathbb{R}$ *is* separable in the standard topology (the rationals $\mathbb{Q} \subset \mathbb{R}$ are dense).

1.1.2. Base of a topology. In practice, it may be awkward to list *all* the open sets constituting a topology; fortunately, one can often define the topology by describing a much smaller collection, which in a sense generates the entire topology.

DEFINITION 1.1.9. A *base* for the topology $\mathcal{T}$ is a subcollection $\beta \subset \mathcal{T}$ such that for any $O \in \mathcal{T}$ there is a $B \in \beta$ for which we have $x \in B \subset O$.

Most topological spaces considered in analysis and geometry (but not in algebraic geometry) have a *countable base.* Such topological spaces are often called *second countable*.

A *base of neighborhoods of a point* x is a collection $\mathcal{B}$ of open neighborhoods of x such that any neighborhood of x contains an element of $\mathcal{B}$. If any point of a topological space has a countable base of neighborhoods, then the space (or the topology) is called *first countable*.

EXAMPLE 1.1.10. Euclidean space $\mathbb{R}^n$ with the standard topology (the usual open and closed sets) has bases consisting of all open balls, open balls of rational radius, open balls of rational center and radius. The latter is a countable base.

EXAMPLE 1.1.11. The real line (or any uncountable set) in the discrete topology (all sets are open) is an example of a first countable but not second countable topological space.

PROPOSITION 1.1.12. *Every topological space with a countable space is separable.*

PROOF. Pick a point in each element of a countable base. The resulting set is at most countable. It is dense since otherwise the complement to its closure would contain an element of the base. □

1.1.3. Comparison of topologies. A topology $\mathcal{S}$ is said to be *stronger* (or *finer*) than $\mathcal{T}$ if $\mathcal{T} \subset \mathcal{S}$, and *weaker* (or *coarser*) if $\mathcal{S} \subset \mathcal{T}$.

There are two extreme topologies on any set: the weakest *trivial topology* with only the whole space and the empty set being open, and the strongest or finest *discrete topology* where all sets are open (and hence closed).

EXAMPLE 1.1.13. On the two point set D, the topology obtained by declaring open (besides D and $\varnothing$) the set consisting of one of the points (but not the other) is strictly finer than the trivial topology and strictly weaker than the discrete topology.

PROPOSITION 1.1.14. *For any set X and any collection $\mathcal{C}$ of subsets of X there exists a unique weakest topology for which all sets from $\mathcal{C}$ are open.*

PROOF. Consider the collection $\mathcal{T}$ which consist of unions of finite intersections of sets from $\mathcal{C}$ and also includes the whole space and the empty set. By properties (2) and (3) of Definition 1.1.1 in any topology in which sets from $\mathcal{C}$ are open the sets from $\mathcal{T}$ are also open. Collection $\mathcal{T}$ satisfies property (1) of Definition 1.1.1 by definition, and it follows immediately from the properties of unions and intersections that $\mathcal{T}$ satisfies (2) and (3) of Definition 1.1.1. □

Any topology weaker than a separable topology is also separable, since any dense set in a stronger topology is also dense in a weaker one.

EXERCISE 1.1.2. How many topologies are there on the 2–element set and on the 3–element set?

EXERCISE 1.1.3. On the integers $\mathbb{Z}$, consider the *profinite* topology for which open sets are defined as unions (not necessarily finite) of arithmetic progressions (non-constant and infinite in both directions). Prove that this defines a topology which is neither discrete nor trivial.

EXERCISE 1.1.4. Define *Zariski* topology in the set of real numbers by declaring complements of finite sets to be open. Prove that this defines a topology which is coarser than the standard one. Give an example of a sequence such that all points are its limits.

EXERCISE 1.1.5. On the set $\mathbb{R} \cup \{*\}$, define a topology by declaring open all sets of the form $\{*\} \cup G$, where $G \subset \mathbb{R}$ is open in the standard topology of $\mathbb{R}$.

(a) Show that this is indeed a topology, coarser than the discrete topology on this set.

(b) Give an example of a convergent sequence which has two limits.

1.2. Continuous maps and homeomorphisms

In this section, we study, in the language of topology, the fundamental notion of continuity and define the main equivalence relation between topological spaces – homeomorphism. We can say (in the category theory language) that now, since the objects (topological spaces) have been defined, we are ready to define the corresponding morphisms (continuous maps) and isomorphisms (topological equivalence or homeomorphism).

1.2.1. Continuous maps. The topological definition of continuity is simpler and more natural than the ε, δ definition familiar from the elementary real analysis course.

DEFINITION 1.2.1. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. A map $f\colon X \to Y$ is said to be *continuous* if $O \in \mathcal{S}$ implies $f^{-1}(O) \in \mathcal{T}$ (preimages of open sets are open):

f is an *open map* if it is continuous and $O \in \mathcal{T}$ implies $f(O) \in \mathcal{S}$ (images of open sets are open);

f is *continuous at the point* x if for any neigborhood A of $f(x)$ in Y the preimage $f^{-1}(A)$ contains a neighborhood of x.

A function f from a topological space to $\mathbb{R}$ is said to be *upper semicontinuous* if $f^{-1}(-\infty, c) \in \mathcal{T}$ for all $c \in \mathbb{R}$:
lower semicontinuous if $f^{-1}(c, \infty) \in \mathcal{T}$ for $c \in \mathbb{R}$.

EXERCISE 1.2.1. Prove that a map is continuous if and only if it is continuous at every point.

Let Y be a topological space. For any collection $\mathcal{F}$ of maps from a set X (without a topology) to Y there exists a unique weakest topology on X which makes all maps from $\mathcal{F}$ continuous; this is exactly the weakest topology with respect to which preimages of all open sets in Y under the maps from $\mathcal{F}$ are open. If $\mathcal{F}$ consists of a single map f, this topology is sometimes called the *pullback topology* on X under the map f.

EXERCISE 1.2.2. Let p be the orthogonal projection of the square K on one of its sides. Describe the pullback topology on K. Will an open (in the usual sense) disk inside K be an open set in this topology?

1.2.2. Topological equivalence. Just as algebraists study groups up to isomorphism or matrices up to a linear conjugacy, topologists study (topological) spaces up to homeomorphism.

DEFINITION 1.2.2. A map $f : X \to Y$ between topological spaces is a *homeomorphism* if it is continuous and bijective with continuous inverse.

If there is a homeomorphism $X \to Y$, then X and Y are said to be *homeomorphic* or sometimes *topologically equivalent*.

A property of a topological space that is the same for any two homeomorphic spaces is said to be a *topological invariant* .

The relation of being homeomorphic is obviously an equivalence relation (in the technical sense: it is reflexive, symmetric, and transitive). Thus topological spaces split into equivalence classes, sometimes called *homeomorphy classes*. In

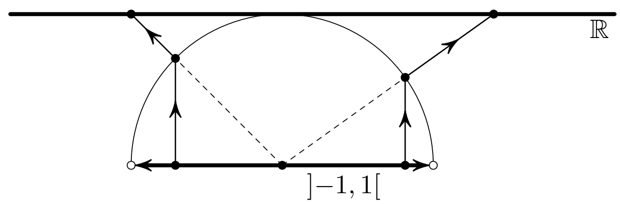

FIGURE 1.2.1. The open interval is homeomorphic to the real line

this connection, the topologist is sometimes described as a person who cannot distinguish a coffee cup from a doughnut (since these two objects are homeomorphic). In other words, two homeomorphic topological spaces are identical or indistinguishable from the intrinsic point of view in the same sense as isomorphic groups are indistinguishable from the point of view of abstract group theory or two conjugate $n \times n$ matrices are indistinguishable as linear transformations of an n-dimensional vector space without a fixed basis.

EXAMPLE 1.2.3. The figure shows how to construct homeomorphisms between the open interval and the open half-circle and between the open half-circle and the real line $\mathbb{R}$, thus establishing that the open interval is homeomorphic to the real line.

EXERCISE 1.2.3. Prove that the sphere $\mathbb{S}^2$ with one point removed is homeomorphic to the plane $\mathbb{R}^2$.

EXERCISE 1.2.4. Prove that any open ball is homeomorphic to $\mathbb{R}^3$.

EXERCISE 1.2.5. Describe a topology on the set $\mathbb{R}^2 \cup \{*\}$ which will make it homeomorphic to the sphere $\mathbb{S}^2$.

To show that certain spaces are homeomorphic one needs to exhibit a homeomorphism; the exercises above give basic but important examples of homeomorphic spaces; we will see many more examples already in the course of this chapter. On the other hand, in order to show that topological spaces are not homeomorphic one need to find an invariant which distinguishes them. Let us consider a very basic example which can be treated with tools from elementary real analysis.

EXAMPLE 1.2.4. In order to show that closed interval is not homeomorphic to an open interval (and hence by Example 1.2.3 to the real line) notice the following. Both closed and open interval as topological spaces have the property that the only sets which are open and closed at the same time are the space itself and the empty set. This follows from characterization of open subsets on the line as finite or countable unions of disjoint open intervals and the corresponding characterization of open subsets of a closed interval as unions of open intervals and semi-open intervals containing endpoints. Now if one takes any point away from an open interval the resulting space with induced topology (see below) will have two proper subsets which are open and closed simultaneously while in the closed (or semi-open) interval removing an endpoint leaves the space which still has no non-trivial subsets which are closed and open.

In Section 1.6 we will develop some of the ideas which appeared in this simple argument systematically.

The same argument can be used to show that the real line $\mathbb{R}$ is not homeomorphic to Euclidean space $\mathbb{R}^n$ for $n \geq 2$ (see Exercise 1.9.7). It is not sufficient however for proving that $\mathbb{R}^2$ is not homeomorphic $\mathbb{R}^3$. Nevertheless, we feel that we intuitively understand the basic structure of the space $\mathbb{R}^n$ and that topological spaces which locally look like $\mathbb{R}^n$ (they are called (n-dimensional) *topological manifolds*) are natural objects of study in topology. Various examples of topological manifolds will appear in the course of this chapter and in **??** we will introduce precise definitions and deduce some basic properties of topological manifolds.

1.3. Basic constructions

1.3.1. Induced topology. If $Y \subset X$, then Y can be made into a topological space in a natural way by taking the *induced topology*

$$\mathcal{T}_Y := \{O \cap Y \mid O \in \mathcal{T}\}.$$

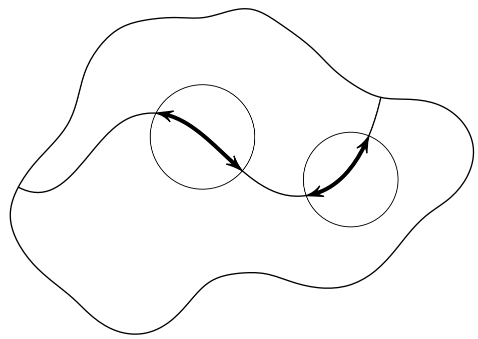

FIGURE 1.3.1. Induced topology

EXAMPLE 1.3.1. The topology induced from $\mathbb{R}^{n+1}$ on the subset

$$\{(x_1, \dots, x_n, x_{n+1}) : \sum_{i=1}^{n+1} x_i^2 = 1\}$$

produces the (standard, or unit) n*–sphere* $\mathbb{S}^n$. For $n = 1$ it is called the (*unit*) *circle* and is sometimes also denoted by $\mathbb{T}$.

EXERCISE 1.3.1. Prove that the boundary of the square is homeomorphic to the circle.

EXERCISE 1.3.2. Prove that the sphere $\mathbb{S}^2$ with any two points removed is homeomorphic to the infinite cylinder $C := \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 = 1\}$.

EXERCISE 1.3.3. Let $S := \{(x, y, z) \in \mathbb{R}^3 \mid z = 0,\ x^2 + y^2 = 1\}$. Show that $\mathbb{R}^3 \setminus S$ can be mapped continuously onto the circle.

1.3.2. Product topology. If $(X_\alpha, \mathcal{T}_\alpha)$, $\alpha \in A$ are topological spaces and A is any set, then the *product topology* on $\prod_{\alpha \in A} X$ is the topology determined by the base

$$\Big\{ \prod_\alpha O_\alpha \mid O_\alpha \in \mathcal{T}_\alpha,\ O_\alpha \neq X_\alpha \text{ for only finitely many } \alpha \Big\}.$$

EXAMPLE 1.3.2. The standard topology in $\mathbb{R}^n$ coincides with the product topology on the product of n copies of the real line $\mathbb{R}$.

EXAMPLE 1.3.3. The product of n copies of the circle is called the n*–torus* and is usually denoted by $\mathbb{T}^n$. The n– torus can be naturally identified with the following subset of $\mathbb{R}^{2n}$:

$$\{(x_1, \dots x_{2n}) : x_{2i-1}^2 + x_{2i}^2 = 1,\ i = 1, \dots, n.\}$$

with the induced topology.

EXAMPLE 1.3.4. The product of countably many copies of the two–point space, each with the discrete topology, is one of the representations of the *Cantor set* (see Section 1.7 for a detailed discussion).

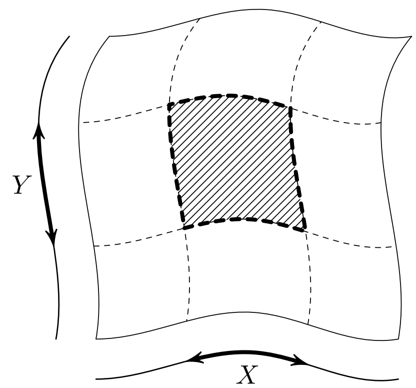

FIGURE 1.3.2. Basis element of the product topology

EXAMPLE 1.3.5. The product of countably many copies of the closed unit interval is called the *Hilbert cube*. It is the first interesting example of a Hausdorff space (Section 1.4) "too big" to lie inside (that is, to be homeomorphic to a subset of) any Euclidean space $\mathbb{R}^n$. Notice however, that not only we lack means of proving the fact right now but the elementary invariants described later in this chapter are not sufficient for this task either.

EXERCISE 1.3.4. Describe a homeomorphism between the Hilbert cube and a closed subset of the unit ball in the Hilbert space l^2 of the square-integrable sequences of reals with topology determined by the norm.

1.3.3. Quotient topology. Consider a topological space $(X, \mathcal{T})$ and suppose there is an equivalence relation $\sim$ defined on X. Let π be the natural projection of X on the set $\hat{X}$ of equivalence classes. The *identification space* or *quotient space* $X/\sim := (\hat{X}, \mathcal{S})$ is the topological space obtained by calling a set $O \subset \hat{X}$ open if $\pi^{-1}(O)$ is open, that is, taking on $\hat{X}$ the finest topology for which π is continuous. For the moment we restrict ourselves to "good" examples, i.e. to the situations where quotient topology is natural in some sense. However the reader should be aware that even very natural equivalence relations often lead to factors with bad properties ranging from the trivial topology to nontrivial ones but lacking basic separation properties (see Section 1.4). We postpone description of such examples till Section 1.8.2.

EXAMPLE 1.3.6. Consider the closed unit interval and the equivalence relation which identifies the endpoints. Other equivalence classes are single points in the interior. The corresponding quotient space is another representation of the circle.

The product of n copies of this quotient space gives another definition of the n–torus.

EXERCISE 1.3.5. Describe the representation of the n–torus from the above example explicitly as the identification space of the unit n–cube I^n:

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n : 0 \le x_i \le 1, \ i = 1, \dots n.$$

EXAMPLE 1.3.7. Consider the following equivalence relation in punctured Euclidean space $\mathbb{R}^{n+1} \setminus \{0\}$:

$$(x_1, \dots, x_{n+1}) \sim (y_1, \dots, y_{n+1}) \text{ iff } y_i = \lambda x_i \text{ for all } i = 1, \dots, n+1$$

with the same real number λ. The corresponding identification space is called the *real projective n–space* and is denoted by $\mathbb{R}P(n)$.

A similar procedure in which λ has to be positive gives another definition of the n–sphere $\mathbb{S}^n$.

EXAMPLE 1.3.8. Consider the equivalence relation in $\mathbb{C}^{n+1} \setminus \{0\}$:

$$(x_1, \dots, x_{n+1}) \sim (y_1, \dots, y_{n+1}) \text{ iff } y_i = \lambda x_i \text{ for all } i = 1, \dots, n+1$$

with the same complex number λ. The corresponding identification space is called the *complex projective n–space* and is detoted by $\mathbb{C}P(n)$.

EXAMPLE 1.3.9. The map $E : [0,1] \to \mathbb{S}^1, E(x) = \exp 2\pi i x$ establishes a homeomorphism between the interval with identified endpoints (Example 1.3.6) and the unit circle defined in Example 1.3.1.

EXAMPLE 1.3.10. The identification of the equator of the 2-sphere to a point yields two spheres with one common point.

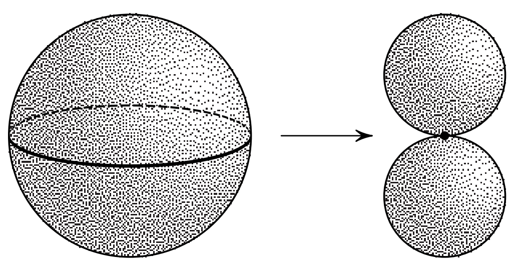

FIGURE 1.3.3. The sphere with equator identified to a point

EXAMPLE 1.3.11. Identifying the short sides of a long rectangle in the natural way yields the lateral surface of the cylinder (which of course is homeomorphic to the annulus), while the identification of the same two sides in the "wrong way" (i.e., after a half twist of the strip) produces the famous Möbius strip. We assume the reader is familiar with the failed experiments of painting the two sides of the Möbius strip in different colors or cutting it into two pieces along its midline. Another less familiar but amusing endeavor is to predict what will happen to the physical object obtained by cutting a paper Möbius strip along its midline if that object is, in its turn, cut along its own midline.

EXERCISE 1.3.6. Describe a homeomorphism between the torus $\mathbb{T}^n$ (Example 1.3.3) and the quotient space described in Example 1.3.6 and the subsequent exercise.

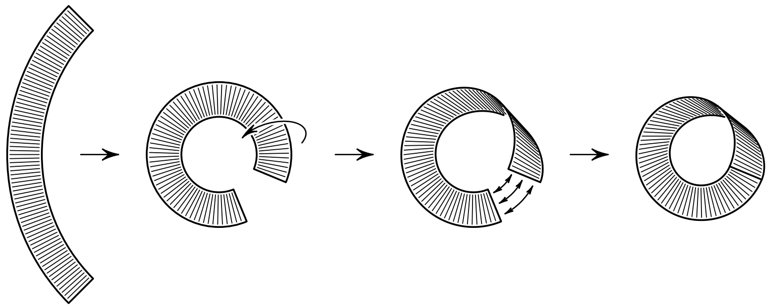

FIGURE 1.3.4. The Möbius strip

EXAMPLE 1.3.12. There are three natural ways to identify points on the pairs of opposite sides of the unit square:

(1) by parallel translations on both pairs; this produces the torus $\mathbb{T}^2$;
(2) by rotations by π around the center of the square; this gives another representation of the projective plane $\mathbb{R}P(2)$
(3) by the parallel translation for one pair and rotation by π for the other; the resulting identification space is called the *Klein bottle*.

EXERCISE 1.3.7. Consider the regular hexagon and identify pairs of opposite sides by corresponding parallel translations. Prove that the resulting identification space is homeomorphic to the torus $\mathbb{T}^2$.

EXERCISE 1.3.8. Describe a homeomorphism between the sphere $\mathbb{S}^n$ (Example 1.3.1) and the second quotient space of Example 1.3.7.

EXERCISE 1.3.9. Prove that the real projective space $\mathbb{R}P(n)$ is homeomorphic to the quotient space of the sphere S^n with respect to the equivalence relation which identifies pairs of opposite points: x and $-x$.

EXERCISE 1.3.10. Consider the equivalence relation on the closed unit ball $\mathbb{D}^n$ in $\mathbb{R}^n$:

$$\{(x_1, \dots, x_n) : \sum_{i=1}^{n} x_i^2 \leq 1\}$$

which identifies all points of $\partial\mathbb{D}^n = \mathbb{S}^{n-1}$ and does nothing to interior points. Prove that the quotient space is homeomorphic to $\mathbb{S}^n$.

EXERCISE 1.3.11. Show that $\mathbb{C}P(1)$ is homeomorphic to $\mathbb{S}^2$.

DEFINITION 1.3.13. The *cone* $\operatorname{Cone}(X)$ over a topological space X is the quotient space obtained by identifying all points of the form $(x, 1)$ in the product $(X \times [0, 1]$ (supplied with the product topology).

The *suspension* $\Sigma(X)$ of a topological space X is the quotient space of the product $X \times [-1, 1]$ obtained by identifying all points of the form $x \times 1$ and identifying all points of the form $x \times -1$. By convention, the suspension of the empty set will be the two-point set $\mathbb{S}^0$.

The *join* $X * Y$ of two topological spaces X and Y , roughly speaking, is obtained by joining all pairs of points (x, y), $x \in X$, $y \in Y$, by line segments and supplying the result with the natural topology; more precisele, $X * Y$ is the quotient space of the product $X \times [-1, 1] \times Y$ under the following identifications:

$(x, -1, y) \sim (x, -1, y')$ for any $x \in X$ and all $y, y' \in Y$,
$(x, 1, y) \sim (x', 1, y)$ for any $y \in Y$ and all $x, x' \in X$.

EXAMPLE 1.3.14. (a) $\mathrm{Cone}(*) = \mathbb{D}^1$ and $\mathrm{Cone}(\mathbb{D}^{n-1}) = \mathbb{D}^n$ for $n > 1$.
(b) The suspension $\Sigma(\mathbb{S}^n)$ of the n-sphere is the $(n+1)$-sphere $\mathbb{S}^{n+1}$.
(c) The join of two closed intervals is the 3-simplex (see the figure).

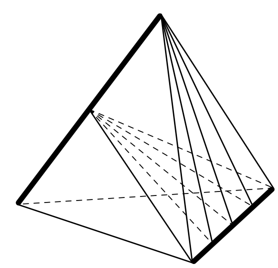

FIGURE 1.3.5. The 3-simplex as the join of two segments

EXERCISE 1.3.12. Show that the cone over the sphere $\mathbb{S}^n$ is (homeomorphic to) the disk $\mathbb{D}^{n+1}$.

EXERCISE 1.3.13. Show that the join of two spheres $\mathbb{S}^k$ and $\mathbb{S}^l$ is (homeomorphic to) the sphere $\mathbb{S}^{k+l+1}$.

EXERCISE 1.3.14. Is the join operation on topological spaces associative?

1.4. Separation properties

Separation properties provide one of the approaches to measuring how fine is a given topology.

1.4.1. T1, Hausdorff, and normal spaces. Here we list, in decreasing order of generality, the most common separation axioms of topological spaces.

DEFINITION 1.4.1. A topological space $(X, \mathcal{T})$ is said to be a
(T1) *space* if any point is a closed set. Equivalently, for any pair of points $x_1, x_2 \in X$ there exists a neighborhood of x_1 not containing x_2;
(T2) or *Hausdorff space* if any two distinct points possess nonintersecting neighborhoods;
(T4) or *normal space* if it is Hausdorff and any two closed disjoint subsets possess nonintersecting neighborhoods. [1]

It follows immediately from the definition of induced topology that any of the above separation properties is inherited by the induced topology on any subset.

[1]Hausdorff (or (T1)) assumption is needed to ensure that there are enough closed sets; specifically that points are closed sets. Otherwise trivial topology would satisfy this property.

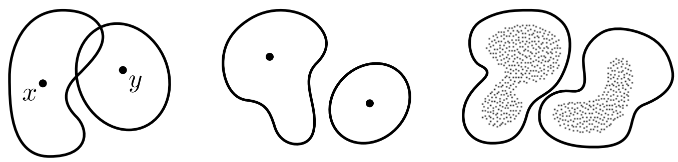

FIGURE 1.4.1. Separation properties

EXERCISE 1.4.1. Prove that in a (T2) space any sequence has no more than one limit. Show that without the (T2) condition this is no longer true.

EXERCISE 1.4.2. Prove that the product of two (T1) (respectively Hausdorff) spaces is a (T1) (resp. Hausdorff) space.

REMARK 1.4.2. We will see later (Section 1.8) that even very naturally defined equivalence relations in nice spaces may produce quotient spaces with widely varying separation properties.

The word "normal" may be understood in its everyday sense like "commonplace" as in "a normal person". Indeed, normal topological possess many properties which one would expect form commonplaces notions of continuity. Here is an examples of such property dealing with extension of maps:

THEOREM 1.4.3. [Tietze] *If X is a normal topological space, $Y \subset X$ is closed, and $f\colon Y \to [-1,1]$ is continuous, then there is a continuous extension of f to X, i.e., a continuous map $F : X \to [-1,1]$ such that $F|_Y = f$.*

The proof is based on the following fundamental result, traditionally called Urysohn Lemma, which asserts existence of many continuous maps from a normal space to the real line and thus provided a basis for introducing measurements in normal topological spaces (see Theorem 2.1.3) and hence by Corollary 2.6.2 also in compact Hausdorff spaces.

THEOREM 1.4.4. [Urysohn Lemma] *If X is a normal topological space and A, B are closed subsets of X, then there exists a continuous map $u : X \to [0,1]$ such that $u(A) = \{0\}$ and $u(B) = \{1\}$.*

PROOF. Let V be en open subset of X and U any subset of X such that $\overline{U} \subset V$. Then there exists an open set W for which $\overline{U} \subset W \subset \overline{W} \subset V$. Indeed, for W we can take any open set containing $\overline{U}$ and not intersecting an open neighborhood of $X \setminus V$ (such a W exists because X is normal).

Applying this to the sets $U := A$ and $V := X \setminus B$, we obtain an "intermediate" open set A_1 such that

$$A \subset A_1 \subset X \setminus B, \tag{1.4.1}$$

where $\overline{A_1} \subset X \setminus B$. Then we can introduce the next intermediate open sets A_1' and A_2 so as to have

$$A \subset A_1' \subset A_1 \subset A_2 \subset X \setminus B, \tag{1.4.2}$$

where each set is contained, together with its closure, in the next one.

For the sequence (1.4.1), we define a function $u_1 : X \to [0,1]$ by setting

$$u_1(x) = \begin{cases} 0 & \text{for } x \in A\ , \\ 1/2 & \text{for } x \in A_1 \setminus A, \\ 1 & \text{for } X \setminus A_1. \end{cases}$$

For the sequence (1.4.2), we define a function $u_2 : X \to [0,1]$ by setting

$$u_2(x) = \begin{cases} 0 & \text{for } x \in A\ , \\ 1/4' & \text{for } x \in A_1' \setminus A\ , \\ 1/2 & \text{for } x \in A_1 \setminus A_1', \\ 3/4 & \text{for } x \in A_2 \setminus A_1, \\ 1 & \text{for } x \in X \setminus A_2. \end{cases}$$

Then we construct a third sequence by inserting intermediate open sets in the sequence (1.4.2) and define a similar function u_3 for this sequence, and so on.

Obviously, $u_2(x) \geq u_1(x)$ for all $x \in X$. Similarly, for any $n > 1$ we have $u_{n+1}(x) \geq u_n(x)$ for all $x \in X$, and therefore the limit function $u(x) := \lim_{n\to infty} u_n(x)$ exists. It only remains to prove that u is continuous.

Suppose that at the nth step we have constructed the nested sequence of sets corresponding to the function u_n

$$A \subset A_1 \subset \ldots A_r \subset X \setminus B,$$

where $\overline{A_i} \subset A_{i+1}$. Let $A_0 := \operatorname{int} A$ be the interior of A, let $A_{-1} := \varnothing$, and $A_{r+1} := X$. Consider the open sets $A_{i+1} \setminus \overline{A_{i-1}}$, $i = 0, 1, \ldots, r$. Clearly,

$$X = \bigcup_{i=0}^{r} (\bar{A}_i \setminus \overline{A_{i-1}}) \subset \bigcup_{i=0}^{r} (A_{i+1} \setminus \overline{A_{i-1}}),$$

so that the open sets $A_{i+1} \setminus \overline{A_{i-1}}$ cover the entire space X.

On each set $A_{i+1} \setminus \overline{A_{i-1}}$ the function takes two values that differ by $1/2^n$. Obviously,

$$|u(x) - u_n(x)| \leq \sum_{k=n+1}^{\infty} 1/2^k \;=\; 1/2^n.$$

For each point $x \in X$ let us choose an open neighborhood of the form $A_{i+1} \setminus \overline{A_{i-1}}$. The image of the open set $A_{i+1} \setminus \overline{A_{i-1}}$ is contained in the interval $(u(x) - \varepsilon, u(x) + \varepsilon)$, where $\varepsilon < 1/2^n$. Taking $\varepsilon \to \infty$, we see that u is continuous. □

Now let us deduce Theorem 1.4.3 from the Urysohn lemma.

To this end, we put

$$r_k := \frac{1}{2}\Big(\frac{2}{3}\Big)^k, \quad k = 1, 2, \dots .$$

Let us construct a sequence of functions $f_1, f_2, \dots$ on X and a sequence of functions $g_1, g_2, \dots$ on Y by induction. First, we put $f_1 := f$. Suppose that the functions $f_1, \dots, f_k$ have been constructed. Consider the two closed disjoint sets

$$A_k := \{x \in X \mid f_k(x) \le -r_k\} \quad \text{and} \quad B_k := \{x \in X \mid f_k(x) \ge r_k\}.$$

Applying the Urysohn lemma to these sets, we obtain a continuous map $g_k : Y \to [-r_k, r_k]$ for which $g_k(A_k) = \{-r_k\}$ and $g_k(B_k) = \{r_k\}$. On the set A_k, the functions f_k and g_k take values in the interval $]-3r_k, -r_k[$; on the set A_k, they take values in the interval $]r_k, 3r_k[$; at all other points of the set X, these functions take values in the interval $]-r_k, r_k[$.

Now let us put $f_{k+1} := f_k - g_k|_X$. The function f_{k+1} is obviously continuous on X and $|f_{k+1}(x)| \le 2r_k = 3r_{k+1}$ for all $x \in X$.

Consider the sequence of functions $g_1, g_2, \dots$ on Y. By construction, $|g_k(y)| \le r_k$ for all $y \in Y$. The series

$$\sum_{k=1}^{\infty} r_k = \frac{1}{2}\sum_{k=1}^{\infty}\Big(\frac{2}{3}\Big)^k$$

converges, and so the series $\Sigma_{k=1}^{\infty} g_k(x)$ converges uniformly on Y to some continuous function

$$F(x) := \sum_{k=1}^{\infty} g_k(x).$$

Further, we have

$$(g_1+\dots+g_k) = (f_1-f_2)+(f_2-f_3)+\dots+(f_k-f_{k+1}) = f_1-f_{k+1} = f-f_{k+1}.$$

But $\lim_{k\to\infty} f_{k+1}(y) = 0$ for any $y \in Y$, hence $F(x) = f(x)$ for any $x \in X$, so that F is a continuous extension of f.

It remains to show that $|F(x)| \le 1$. We have

$$\begin{aligned} |F(x)| &\le \sum_{k=1}^{\infty} |g_k(x)| \le \sum_{k=1}^{\infty} r_k = \sum_{k=1}^{\infty}\Big(\frac{2}{3}\Big)^k \\ &= \sum_{k=1}^{\infty}\Big(\frac{2}{3}\Big)^k = \frac{1}{3}\Big(1-\frac{2}{3}\Big)^{-1} = 1. \qquad \square \end{aligned}$$

COROLLARY 1.4.5. *Let $X \subset Y$ be a closed subset of a normal space Y and let $f : X \to \mathbb{R}$ be continuous. Then f has a continuous extension $F : Y \to \mathbb{R}$.*

PROOF. The statement follows from the Tietze theorem and the Urysohn lemma by appropriately using the rescaling homeomorphism

$$g : \mathbb{R} \to (-\pi/2, \pi/2) \quad \text{given by} \quad g(x) := \arctan(x). \quad \square$$

Most natural topological spaces which appear in analysis and geometry (but not in some branches of algebra) are normal. The most important instance of non-normal topology is discussed in the next subsection.

1.4.2. Zariski topology. The topology that we will now introduce and seems pathological in several aspects (it is non-Hausdorff and does not possess a countable base), but very useful in applications, in particular in algebraic geometry. We begin with the simplest case which was already mentioned in Example 1.1.4

DEFINITION 1.4.6. The *Zariski topology* on the real line $\mathbb{R}$ is defined as the family $\mathcal{Z}$ of all complements to finite sets.

PROPOSITION 1.4.7. *The Zariski topology given above endows* $\mathbb{R}$ *with the structure of a topological space* $(\mathbb{R}, \mathcal{Z})$*, which possesses the following properties:*

(1) *it is a* (T1) *space;*
(2) *it is separable;*
(3) *it is not a Hausdorff space;*
(4) *it does not have a countable base.*

PROOF. All four assertions are fairly straightforward:

(1) the Zariski topology on the real line is (T1), because the complement to any point is open;

(2) it is separable, since it is weaker than the standard topology in $\mathbb{R}$;

(3) it is not Hausdorff, because any two nonempty open sets have nonempty intersection;

(4) it does not have a countable base, because the intersection of all the sets in any countable collection of open sets is nonemply and thus the complement to any point in that intersection does not contain any element from that collection. □

The definition of Zariski topology on $\mathbb{R}$ (Definition 1.4.6) can be straightforwardly generalized to $\mathbb{R}^n$ for any $n \geq 2$, and the assertions of the proposition above remain true. However, this definition is not the natural one, because it generalizes the "wrong form" of the notion of Zariski topology. The "correct form" of that notion originally appeared in algebraic geometry (which studies zero sets of polynomials) and simply says that closed sets in the Zariski topology on $\mathbb{R}$ are sets of zeros of polynomials $p(x) \in \mathbb{R}[x]$. This motivates the following definitions.

DEFINITION 1.4.8. The *Zariski topology* is defined

- in Euclidean space $\mathbb{R}^n$ by stipulating that the sets of zeros of all polynomials are closed;
- on the real and complex projective spaces $\mathbb{R}P(n)$ and $\mathbb{C}P(n)$ (Example 1.3.7, Example 1.3.8) via zero sets of homogeneous polynomials in $n+1$ real and complex variables respectively.

EXERCISE 1.4.3. Verify that the above definitions supply each of the sets $\mathbb{R}^n$, $\mathbb{R}P(n)$, and $\mathbb{C}P(n)$ with the structure of a topological space satisfying the assertions of Proposition 1.4.7.

1.5. Compactness

The fundamental notion of compactness, familiar from the elementary real analysis course for subsets of the real line $\mathbb{R}$ or of Euclidean space $\mathbb{R}^n$, is defined below in the most general topological situation.

1.5.1. Types of compactness. A family of open sets $\{O_\alpha\} \subset \mathcal{T}$, $\alpha \in A$ is called an *open cover* of a topological space X if $X = \bigcup_{\alpha\in A} O_\alpha$, and is a finite open cover if A is finite.

DEFINITION 1.5.1. The space $(X, \mathcal{T})$ is called
• *compact* if every open cover of X has a finite subcover;
• *sequentially compact* if every sequence has a convergent subsequence;
• *σ–compact* if it is the union of a countable family of compact sets.
• *locally compact* if every point has an open neighborhood whose closure is compact in the induced topology.

It is known from elementary real analysis that for subsets of a $\mathbb{R}^n$ compactness and sequential compactness are equivalent. This fact naturally generalizes to metric spaces (see Proposition 2.7.4).

PROPOSITION 1.5.2. *Any closed subset of a compact set is compact.*

PROOF. If K is compact, $C \subset K$ is closed, and Γ is an open cover for C, then $\Gamma_0 := \Gamma \cup \{K \smallsetminus C\}$ is an open cover for K, hence Γ_0 contains a finite subcover $\Gamma' \cup \{K \smallsetminus C\}$ for K; therefore Γ' is a finite subcover (of Γ) for C. □

PROPOSITION 1.5.3. *Any compact subset of a Hausdorff space is closed.*

PROOF. Let X be Hausdorff and let $C \subset X$ be compact. Fix a point $x \in X \smallsetminus C$ and for each $y \in C$ take neighborhoods U_y of y and V_y of x such that $U_y \cap V_y = \varnothing$. Then $\bigcup_{y\in C} U_y \supset C$ is a cover of C and has a finite subcover $\{U_{x_i} \mid 0 \le i \le n\}$. Hence $N_x := \bigcap_{i=0}^n V_{y_i}$ is a neighborhood of x disjoint from C. Thus

$$X \smallsetminus C = \bigcup_{x\in X\smallsetminus C} N_x$$

is open and therefore C is closed. □

PROPOSITION 1.5.4. *Any compact Hausdorff space is normal.*

PROOF. First we show that a closed set K and a point $p \notin K$ can be separated by open sets. For $x \in K$ there are open sets O_x, U_x such that $x \in O_x$, $p \in U_x$ and $O_x \cap U_x = \varnothing$. Since K is compact, there is a finite subcover $O := \bigcup_{i=1}^n O_{x_i} \supset K$, and $U := \bigcap_{i=1}^n U_{x_i}$ is an open set containing p disjoint from O.

Now suppose K, L are closed sets. For $p \in L$, consider open disjoint sets $O_p \supset K$, $U_p \ni p$. By the compactness of L, there is a finite subcover $U := \bigcup_{j=1}^m U_{p_j} \supset L$, and so $O := \bigcap_{j=1}^m O_{p_j} \supset K$ is an open set disjoint from $U \supset L$. □

DEFINITION 1.5.5. A collection of sets is said to have the *finite intersection property* if every finite subcollection has nonempty intersection.

PROPOSITION 1.5.6. *Any collection of compact sets with the finite intersection property has a nonempty intersection.*

PROOF. It suffices to show that in a compact space every collection of closed sets with the finite intersection property has nonempty intersection. Arguing by contradiction, suppose there is a collection of closed subsets in a compact space K with empty intersection. Then their complements form an open cover of K. Since it has a finite subcover, the finite intersection property does not hold. $\square$

EXERCISE 1.5.1. Show that if the compactness assumption in the previous proposition is omitted, then its assertion is no longer true.

EXERCISE 1.5.2. Prove that a subset of $\mathbb{R}$ or of $\mathbb{R}^n$ is compact iff it is closed and bounded.

1.5.2. Compactifications of non-compact spaces.

DEFINITION 1.5.7. A compact topological space K is called a *compactification* of a Hausdorff space $(X, \mathcal{T})$ if K contains a dense subset homeomorphic to X.

The simplest example of compactification is the following.

DEFINITION 1.5.8. The *one-point compactification* of a noncompact Hausdorff space $(X, \mathcal{T})$ is $\hat{X} := (X \cup \{\infty\}, \mathcal{S})$, where

$$\mathcal{S} := \mathcal{T} \cup \{(X \cup \{\infty\}) \smallsetminus K \mid K \subset X \text{ compact}\}.$$

EXERCISE 1.5.3. Show that the one-point compactification of a Hausdorff space X is a compact (T1) space with X as a dense subset. Find a necessary and sufficient condition on X which makes the one-point compactification Hausdorff.

EXERCISE 1.5.4. Describe the one-point compactification of $\mathbb{R}^n$.

Other compactifications are even more important.

EXAMPLE 1.5.9. Real projective space $\mathbb{R}P(n)$ is a compactification of the Euclidean space $\mathbb{R}^n$. This follows easily form the description of $\mathbb{R}P(n)$ as the identification space of a (say, northern) hemisphere with pairs of opposite equatorial points identified. The open hemisphere is homeomorphic to $\mathbb{R}^n$ and the attached "set at infinity" is homeomorphic to the projective space $\mathbb{R}P(n-1)$.

EXERCISE 1.5.5. Describe the complex projective space $\mathbb{C}P(n)$ (see Example 1.3.8) as a compactification of the space $\mathbb{C}^n$ (which is of course homeomorphic to $\mathbb{R}^{2n}$). Specifically, identify the set of added "points at infinity" as a topological space. and desribe open sets which contain points at infinity.

1.5.3. Compactness under products, maps, and bijections. The following result has numerous applications in analysis, PDE, and other mathematical disciplines.

THEOREM 1.5.10. *The product of any family of compact spaces is compact.*

PROOF. Consider an open cover $\mathcal{C}$ of the product of two compact topological spaces X and Y. Since any open neighborhood of any point contains the product of opens subsets in x and Y we can assume that every element of $\mathcal{C}$ is the product of open subsets in X and Y. Since for each $x \in X$ the subset $\{x\} \times Y$ in the induced topology is homeomorphic to Y and hence compact, one can find a finite subcollection $\mathcal{C}_x \subset \mathcal{C}$ which covers $\{x\} \times Y$.

For $(x, y) \in X \times Y$, denote by p_1 the projection on the first factor: $p_1(x, y) = x$. Let $U_x = \bigcap_{C \in \mathcal{O}_x} p_1(C)$; this is an open neighborhood of x and since the elements of $\mathcal{O}_x$ are products, $\mathcal{O}_x$ covers $U_x \times Y$. The sets $\mathcal{U}_x$, $x \in X$ form an open cover of X. By the compactness of X, there is a finite subcover, say $\{U_{x_1}, \ldots, U_{x_k}\}$. Then the union of collections $\mathcal{O}_{x_1}, \ldots, \mathcal{O}_{x_k}$ form a finite open cover of $X \times Y$.

For a finite number of factors, the theorem follows by induction from the associativity of the product operation and the case of two factors. The proof for an arbitrary number of factors uses some general set theory tools based on axiom of choice. □

PROPOSITION 1.5.11. *The image of a compact set under a continuous map is compact.*

PROOF. If C is compact and $f \colon C \to Y$ continuous and surjective, then any open cover Γ of Y induces an open cover $f_*\Gamma := \{f^{-1}(O) \mid O \in \Gamma\}$ of C which by compactness has a finite subcover $\{f^{-1}(O_i) \mid i = 1, \ldots, n\}$. By surjectivity, $\{O_i\}_{i=1}^n$ is a cover for Y. □

Since the real line is an ordered set and any compact subset is bounded and contains the maximal and the minimal element we immediately obtain an important classical result from real analysis.

COROLLARY 1.5.12. *Any continuous real-valued function on a compact topological space is bounded from above and below and attains its maximal and minimal values.*

A useful application of the notions of continuity, compactness, and separation is the following simple but fundamental result, sometimes referred to as *invariance of domain*:

PROPOSITION 1.5.13. *A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.*

PROOF. Suppose X is compact, Y Hausdorff, $f \colon X \to Y$ bijective and continuous, and $O \subset X$ open. Then $C := X \smallsetminus O$ is closed, hence compact, and $f(C)$ is compact, hence closed, so $f(O) = Y \smallsetminus f(C)$ (by bijectivity) is open. □

Using Proposition 1.5.4 we obtain

COROLLARY 1.5.14. *Under the assumption of Proposition 1.5.13 spaces X and Y are normal.*

EXERCISE 1.5.6. Show that for noncompact X the assertion of Proposition 1.5.13 no longer holds.

1.6. Connectedness and path connectedness

There are two rival formal definitions of the intuitive notion of connectedness of a topological space. The first is based on the idea that such a space "consists of one piece" (i.e., does not "fall apart into two pieces"), the second interprets connectedness as the possibility of "moving continuously from any point to any other point".

1.6.1. Definition and invariance under continuous maps.

DEFINITION 1.6.1. A topological space $(X, \mathcal{T})$ is said to be

• *connected* if X cannot be represented as the union of two nonempty disjoint open sets (or, equivalently, two nonempty disjoint closed sets);

• *path connected* if for any two points $x_0, x_1 \in X$ there exists a path joining x_0 to x_1, i.e., a continuous map $c\colon [0,1] \to X$ such that $c(i) = x_i,\ i = \{0,1\}$.

PROPOSITION 1.6.2. *The continuous image of a connected space X is connected.*

PROOF. If the image is decomposed into the union of two disjoint open sets, the preimages of theses sets which are open by continuity would give a similar decomposition for X. □

PROPOSITION 1.6.3. (1) *Interval is connected*
(2) *Any path-connected space is connected.*

PROOF. (1) Any open subset X of an interval is the union of disjoint open subintervals. The complement of X contains the endpoints of those intervals and hence cannot be open.

(2) Suppose X is path-connected and let $x = X_0 \cup X_1$, where X_0 and X_1 are open and nonempty. Let $x_0 \in X_0,\ x_1 \in X_1$ and $c\colon [0,1] \to X$ is a continuous map such that $c(i) = x_i,\ i \in \{0,1\}$. By Proposition 1.6.2 the image $c([0,1])$ is a connected subset of X in induced topology which is decomposed into the union of two nonempty open subsets $c([0,1]) \cap X_0$ and $c([0,1]) \cap X_1$, a contradiction. □

REMARK 1.6.4. Connected space may not be path-connected as is shown by the union of the graph of $\sin 1/x$ and $\{0\} \times [-1,1]$ in $\mathbb{R}^2$ (see the figure).

PROPOSITION 1.6.5. *The continuous image of a path connected space X is path connected.*

PROOF. Let $f : X \to Y$ be continuous and surjective; take any two points $y_1, y_2 \in Y$. Then by surjectivity the sets $f^{-1}(y_i)$, $i = 1,2$ are nonempty and we can choose points $x_i \in f^{-1}(y_1)$, $i = 1,2$. Since X is path connected, there is a path $\alpha : [0,1] \to X$ joining x_1 to x_2. But then the path $f \circ \alpha$ joins y_1 to y_2. □

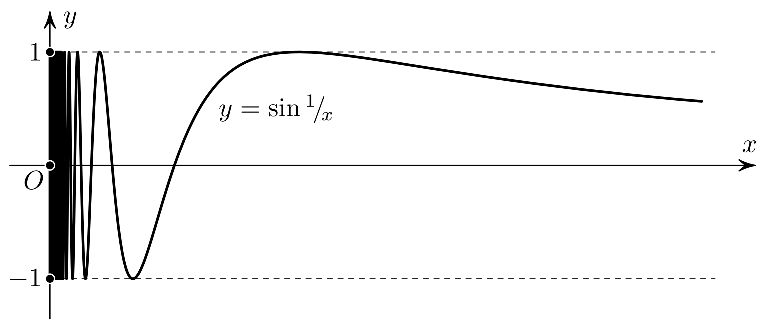

FIGURE 1.6.1. Connected but not path connected space

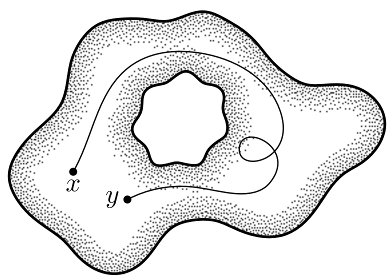

FIGURE 1.6.2. Path connectedness

1.6.2. Products and quotients.

PROPOSITION 1.6.6. *The product of two connected topological spaces is connected.*

PROOF. Suppose X, Y are connected and assume that $X \times Y = A \cup B$, where A and B are open, and $A \cap B = \varnothing$. Then either $A = X_1 \times Y$ for some open $X_1 \subset X$ or there exists an $x \in X$ such that $\{x\} \times Y \cap A \neq \varnothing$ and $\{x\} \times Y \cap B \neq \varnothing$.

The former case is impossible, else we would have $B = (X \setminus X_1) \times Y$ and so $X = X_1 \cup (X \setminus X_1)$ would not be connected.

In the latter case, $Y = p_2(\{x\} \times Y \cap A) \cup p_2(\{x\} \times Y \cap B)$ (where $p_2(x, y) = y$ is the projection on the second factor) that is, $\{x\} \times Y$ is the union of two disjoint open sets, hence not connected. Obviously p_2 restricted to $\{x\} \times Y$ is a homeomorphism onto Y, and so Y is not connected either, a contradiction. □

PROPOSITION 1.6.7. *The product of two path-connected topological spaces is connected.*

PROOF. Let $(x_0, y_0),\ (x_1, y_1) \in X \times Y$ and $c_X,\ c_Y$ are paths connecting x_0 with x_1 and y_0 with y_1 correspondingly. Then the path $c\colon [0, 1] \to X \times Y$ defined by

$$c(t) = (c_X(t), c_Y(t))$$

connects (x_0, y_0) with (x_1, y_1). □

The following property follows immediately from the definition of the quotient topology

PROPOSITION 1.6.8. *Any quotient space of a connected topological space is connected.*

1.6.3. Connected subsets and connected components. A subset of a topological space is *connected* (*path connected*) if it is a connected (path connected) space in the induced topology.

A *connected component* of a topological space X is a maximal connected subset of X.

A *path connected component* of X is a maximal path connected subset of X.

PROPOSITION 1.6.9. *The closure of a connected subset $Y \subset X$ is connected.*

PROOF. If $\bar{Y} = Y_1 \cup Y_2$, where $Y_1,\ Y_2$ are open and $Y_1 \cap Y_2 = \varnothing$, then since the set Y is dense in its closure $Y = (Y \cap Y_1) \cup (Y \cap Y_2)$ with both $Y \cap Y_1$ and $Y \cap Y_1$ open in the induced topology and nonempty. □

COROLLARY 1.6.10. *Connected components are closed.*

PROPOSITION 1.6.11. *The union of two connected subsets $Y_1, Y_2 \subset X$ such that $Y_1 \cap Y_2 \neq \varnothing$, is connected.*

PROOF. We will argue by contradiction. Assume that $Y_1 \cap Y_2$ is the disjoint union of of open sets Z_1 and Z_2. If $Z_1 \supset Y_1$, then $Y_2 = Z_2 \cup (Z_1 \cap Y_2)$ and hence Y_2 is not connected. Similarly, it is impossible that $Z_2 \supset Y_1$. Thus $Y_1 \cap Z_i \neq \varnothing,\ i = 1, 2$ and hence $Y_1 = (Y_1 \cap Z_1) \cup (Y_1 \cap Z_2)$ and hence Y_1 is not connected. □

1.6.4. Decomposition into connected components. For any topological space there is a unique *decomposition into connected components* and a unique *decomposition into path connected components*. The elements of these decompositions are equivalence classes of the following two equivalence relations respectively:

(i) x is equivalent to y if there exists a connected subset $Y \subset X$ which contains x and y.

In order to show that the equivalence classes are indeed connected components, one needs to prove that they are connected. For, if A is an equivalence class, assume that $A = A_1 \cup A_2$, where A_1 and A_2 are disjoint and open. Pick $x_1 \in A_1$ and $x_2 \in A_2$ and find a closed connected set A_3 which contains both points. But then $A \subset (A_1 \cup A_3) \cup A_2$, which is connected by Proposition 1.6.11. Hence $A = (A_1 \cup A_3) \cup A_2)$ and A is connected.

(ii) x is equivalent to y if there exists a continuous curve $c\colon [0, 1] \to X$ with $c(0) = x,\ c(1) = y$

REMARK 1.6.12. The closure of a path connected subset may be fail to be path connected. It is easy to construct such a subset by looking at Remark 1.6.4

1.6.5. Arc connectedness. Arc connectedness is a more restrictive notion than path connectedness: a topological space X is called *arc connected* if, for any two distinct points $x, y \in X$ there exist an arc joining them, i.e., there is an injective continuous map $h : [0,1] \to X$ such that $h(0) = x$ and $h(1) = y$.

It turns out, however, that arc connectedness is not a much more stronger requirement than path connectedness – in fact the two notions coincide for Hausdorff spaces.

THEOREM 1.6.13. *A Hausdorff space is arc connected if and only if it is path connected.*

PROOF. Let X be a path-connected Hausdorff space, $x_0, x_1 \in X$ and $c\colon [0,1] \to X$ a continuous map such that $c(i) = x_i,\ i = 0,1$. Notice that the image $c([0,1])$ is a compact subset of X by Proposition 1.5.11 even though we will not use that directly. We will change the map c within this image by successively cutting of superfluous pieces and rescaling what remains.

Consider the point $c(1/2)$. If it coincides with one of the endpoints x_o or x_1 we define $c_1(t)$ as $c(2t-1)$ or $c(2t)$ correspondingly. Otherwise consider pairs $t_0 < 1/2 < t_1$ such that $c(t_0) = c(t_1)$. The set of all such pairs is closed in the product $[0,1]\times[0,1]$ and the function $|t_0 - t_1|$ reaches maximum on that set. If this maximum is equal to zero the map c is already injective. Otherwise the maximum is positive and is reached at a pair (a_1, b_1). we define the map c_1 as follows

$$c_1(t) = \begin{cases} c(t/2a_1),\ if\ \ 0 \le t \le a_1, \\ c(1/2),\ if\ a_1 \le t \le b_1, \\ c(t/2(1-b_1) + (1-b_1)/2),\ if\, b_1 \le t \le 1. \end{cases}$$

Notice that $c_1([0,1/2))$ and $c_1((1/2,1])$ are disjoint since otherwise there would exist $a' < a_1 < b_1 < b'$ such that $c(a') = c(b')$ contradicting maximality of the pair (a_1, b_1).

Now we proceed by induction. We assume that a continuous map $c_n\colon [0,1] \to c([0,1])$ has been constructed such that the images of intervals $(k/2^n, (k+1)/2^n),\ k = 0,\dots,2^n - 1$ are disjoint. Furthermore, while we do not exclude that $c_n(k/2^n) = c_n((k+1)/2^n)$ we assume that $c_n(k/2^n) \neq c_n(l/2^n)$ if $|k - l| > 1$.

We find a_n^k, b_n^k maximizing the difference $|t_0 - t_1|$ among all pairs

$$(t_0, t_1):\ \ k/2^n \le t_0 \le t_1 \le (k+1)/2^n$$

and construct the map c_{n+1} on each interval $[k/2^n, (k+1)/2^n]$ as above with c_n in place of c and a_n^k, b_n^k in place of a_1, b_1 with the proper renormalization. As before special provision are made if c_n is injective on one of the intervals (in this case we set $c_{n+1} = c_n$) of if the image of the midpoint coincides with that of one of the endpoints (one half is cut off that the other renormalized). □

EXERCISE 1.6.1. Give an example of a path connected but not arc connected topological space.

1.7. Totally disconnected spaces and Cantor sets

On the opposite end from connected spaces are those spaces which do not have any connected nontrivial connected subsets at all.

1.7.1. Examples of totally disconnected spaces.

DEFINITION 1.7.1. A topological space $(X, \mathcal{T})$ is said to be *totally disconnected* if every point is a connected component. In other words, the only connected subsets of a totally disconnected space X are single points.

Discrete topologies (all points are open) give trivial examples of totally disconnected topological spaces. Another example is the set

$$\left\{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots,\right\}$$

with the topology induced from the real line. More complicated examples of compact totally disconnected space in which isolated points are dense can be easily constructed. For instance, one can consider the set of rational numbers $\mathbb{Q} \subset \mathbb{R}$ with the induced topology (which is not locally compact).

The most fundamental (and famous) example of a totally disconnected set is the Cantor set, which we now define.

DEFINITION 1.7.2. The (standard middle-third) *Cantor set* $C(1/3)$ is defined as follows:

$$C(1/3); = \left\{x \in \mathbb{R} : x = \sum_{i=1}^{\infty} \frac{x_i}{3^i},\ x_i \in \{0, 2\},\ i = 1, 2, \dots\right\}.$$

Geometrically, the construction of the set $C(1/3)$ may be described in the following way: we start with the closed interval $[0, 1]$, divide it into three equal subintervals and throw out the (open) middle one, divide each of the two remaining ones into equal subintervals and throw out the open middle ones and continue this process *ad infinitum*. What will be left? Of course the (countable set of) endpoints of the removed intervals will remain, but there will also be a much larger (uncountable) set of remaining "mysterious points", namely those which do not have the ternary digit 1 in their ternary expansion.

1.7.2. Lebesgue measure of Cantor sets. There are many different ways of constructing subsets of $[0, 1]$ which are homeomorphic to the Cantor set $C(1/3)$. For example, instead of throwing out the middle one third intervals at each step, one can do it on the first step and then throw out intervals of length $\frac{1}{18}$ in the middle of two remaining interval and inductively throw out the interval of length $\frac{1}{2^n 3^{n+1}}$ in the middle of each of 2^n intervals which remain after n steps. Let us denote the resulting set $\hat{C}$

EXERCISE 1.7.1. Prove (by computing the infinite sum of lengths of the deleted intervals) that the Cantor set $C(1/3)$ has Lebesgue measure 0 (which was to be expected), whereas the set $\hat{C}$, although nowhere dense, has *positive* Lebesgue measure.

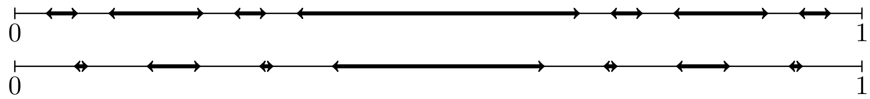

FIGURE 1.7.1. Two Cantor sets

1.7.3. Some other strange properties of Cantor sets. Cantor sets can be obtained not only as subsets of $[0, 1]$, but in many other ways as well.

PROPOSITION 1.7.3. *The countable product of two point spaces with the discrete topology is homeomorphic to the Cantor set.*

PROOF. To see that, identify each factor in the product with $\{0, 2\}$ and consider the map

$$(x_1, x_2, \dots) \mapsto \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \quad x_i \in \{0, 2\},\ i = 1, 2, \dots.$$

This map is a homeomorphism between the product and the Cantor set. □

PROPOSITION 1.7.4. *The product of two* (*and hence of any finite number*) *of Cantor sets is homeomorphic to the Cantor set.*

PROOF. This follows immediately, since the product of two countable products of two point spaces can be presented as such a product by mixing coordinates. □

EXERCISE 1.7.2. Show that the product of countably many copies of the Cantor set is homeomorphic to the Cantor set.

The Cantor set is a compact Hausdorff with countable base (as a closed subset of $[0, 1]$), and it is *perfect* i.e. has no isolated points. As it turns out, it is a universal model for compact totally disconnected perfect Hausdorff topological spaces with countable base, in the sense that any such space is homeomorphic to the Cantor set $C(1/3)$. This statement will be proved later by using the machinery of metric spaces (see Theorem 2.7.9). For now we restrict ourselves to a certain particular case.

PROPOSITION 1.7.5. *Any compact perfect totally disconnected subset A of the real line $\mathbb{R}$ is homeomorphic to the Cantor set.*

PROOF. The set A is bounded, since it is compact, and nowhere dense (does not contain any interval), since it is totally disconnected. Suppose $m = \inf A$ and $M = \sup A$. We will outline a construction of a strictly monotone function $F : [0, 1] \to [m, M]$ such that $F(C) = A$. The set $[m, M] \setminus A$ is the union of countably many disjoint intervals without common ends (since A is perfect). Take one of the intervals whose length is maximal (there are finitely many of them); denote it by I. Define F on the interval I as the increasing linear map whose image is the interval

$[1/3, 2/3]$. Consider the longest intervals I_1 and I_2 to the right and to the left to I. Map them linearly onto $[1/9.2/9]$ and $[7/9, 8/9]$, respectively. The complement $[m, M] \setminus (I_1 \cup I \cup I_2)$ consists of four intervals which are mapped linearly onto the middle third intervals of $[0, 1] \setminus ([1/9.2/9] \cup [1/3, 2/3] \cup [7/9, 8/9]$ and so on by induction. Eventually one obtains a strictly monotone bijective map $[m, M] \setminus A \to [0, 1] \setminus C$ which by continuity is extended to the desired homeomorphism. □

EXERCISE 1.7.3. Prove that the product of countably many finite sets with the discrete topology is homeomorphic to the Cantor set.

1.8. Orbit spaces for group actions

An important class of quotient spaces appears when the equivalence relation is given by the action of a group X by homeomorphisms of a topological space X.

1.8.1. Main definition and nice examples. The notion of a group acting on a space, which formalizes the idea of symmetry, is one of the most important in contemporary mathematics and physics.

DEFINITION 1.8.1. An *action* of a group G on a topological space X is a map $G \times X \to X$, $(g, x) \mapsto xg$ such that

(1) $(xg)h = x(g \cdot h)$ for all $g, h \in G$;

(2) $(x)e = x$ for all $x \in X$, where e is the unit element in G.

The equivalence classes of the corresponding identification are called *orbits* of the action of G on X.

The identification space in this case is denoted by X/G and called the *quotient of X by G* or the *orbit space* of X under the action of G.

We use the notation xg for the point to which the element g takes the point x, which is more convenient than the notation $g(x)$ (nevertheless, the latter is also often used). To specify the chosen notation, one can say that G acts on X *from the right* (for our notation) or *from the left* (when the notation $g(x)$ or gx is used).

Usually, in the definition of an action of a group G on a space X, the group is supplied with a topological structure and the action itself is assumed continuous. Let us make this more precise.

A *topological group* G is defined as a topological Hausdorff space supplied with a continuous group operation, i.e., an operation such that the maps $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous. If G is a finite or countable group, then it is supplied with the discrete topology. When we speak of the action of a topological group G on a space X, we tacitly assume that the map $X \times G \to X$ is a continuous map of topological spaces.

EXAMPLE 1.8.2. Let X be the plane $\mathbb{R}^2$ and G be the rotation group $SO(2)$. Then the orbits are all the circles centered at the origin and the origin itself. The orbit space of $\mathbb{R}^2$ under the action of $SO(2)$ is in a natural bijective correspondence with the half-line $\mathbb{R}_+$.

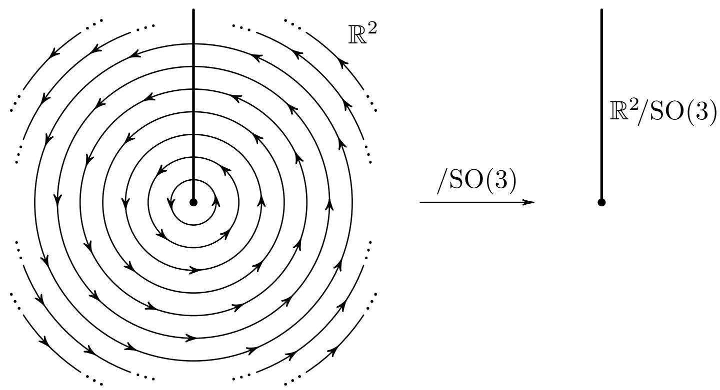

FIGURE 1.8.1. Orbits and identification space of $SO(2)$ action on $\mathbb{R}^2$

The main issue in the present section is that in general the quotient space even for a nice looking group acting on a good (for example, locally compact normal with countable base) topological space may not have good separation properties. The (T1) property for the identification space is easy to ascertain: every orbit of the action must be closed. On the other hand, there does not seem to be a natural necessary and sufficient condition for the quotient space to be Hausdorff. Some useful sufficient conditions will appear in the context of metric spaces.

Still, lots of important spaces appear naturally as such identification spaces.

EXAMPLE 1.8.3. Consider the natural action of the integer lattice $\mathbb{Z}^n$ by translations in $\mathbb{R}^n$. The orbit of a point $p \in \mathbb{R}^n$ is the copy of the integer lattice $\mathbb{Z}^n$ translated by the vector p. The quotient space is homeomorphic to the torus $\mathbb{T}^n$.

An even simpler situation produces a very interesting example.

EXAMPLE 1.8.4. Consider the action of the cyclic group of two elements on the sphere S^n generated by the central symmetry: $Ix = -x$. The corresponding quotient space is naturally identified with the real projective space $\mathbb{R}P(n)$.

EXERCISE 1.8.1. Consider the cyclic group of order q generated by the rotation of the circle by the angle $2\pi/q$. Prove that the identification space is homeomorphic to the circle.

EXERCISE 1.8.2. Consider the cyclic group of order q generated by the rotation of the plane $\mathbb{R}^2$ around the origin by the angle $2\pi/q$. Prove that the identification space is homeomorphic to $\mathbb{R}^2$.

1.8.2. Not so nice examples. Here we will see that even simple actions on familiar spaces can produce unpleasant quotients.

EXAMPLE 1.8.5. Consider the following action A of $\mathbb{R}$ on $\mathbb{R}^2$: for $t \in \mathbb{R}$ let $A_t(x, y) = (x + ty, y)$. The orbit space can be identified with the union of two coordinate axis: every point on the x-axis is fixed and every orbit away from it intersects the y-axis at a single point. However the quotient topology is weaker than the topology induced from $\mathbb{R}^2$ would be. Neighborhoods of the points on the y-axis are ordinary but any neighborhood of a point on the x-axis includes a small open interval of the y-axis around the origin. Thus points on the x-axis cannot be separated by open neighborhoods and the space is (T1) (since orbits are closed) but not Hausdorff.

An even weaker but still nontrivial separation property appears in the following example.

EXAMPLE 1.8.6. Consider the action of $\mathbb{Z}$ on $\mathbb{R}$ generated by the map $x \to 2x$. The quotient space can be identified with the union of the circle and an extra point p. Induced topology on the circle is standard. However, the only open set which contains p is the whole space! See Exercise 1.9.26.

Finally let us point out that if all orbits of an action are dense, then the quotient topology is obviously trivial: there are no invariant open sets other than $\varnothing$ and the whole space. Here is a concrete example.

EXAMPLE 1.8.7. Consider the action T of $\mathbb{Q}$, the additive group of rational number on $\mathbb{R}$ by translations: put $T_r(x) = x + r$ for $r \in \mathbb{Q}$ and $x \in \mathbb{R}$. The orbits are translations of $\mathbb{Q}$, hence dense. Thus the quotient topology is trivial.

1.9. Problems

EXERCISE 1.9.1. How many non-homeomorphic topologies are there on the 2–element set and on the 3–element set?

EXERCISE 1.9.2. Let $S := \{(x, y, z) \in \mathbb{R}^3 \,|\, z = 0,\ x^2 + y^2 = 1\}$. Show that $\mathbb{R}^3 \setminus S$ can be mapped continuously onto the circle.

EXERCISE 1.9.3. Consider the product topology on the product of countably many copies of the real line. (this product space is sometimes denoted $\mathbb{R}^\infty$).

a) Does it have a countable base?

b) Is it separable?

EXERCISE 1.9.4. Consider the space $\mathcal{L}$ of all bounded maps $\mathbb{Z} \to \mathbb{Z}$ with the topology of pointwise convergece.

a) Describe the open sets for this topology.

b) Prove that $\mathcal{L}$ is the countable union of disjoint closed subsets each homeomorphic to a Cantor set.

Hint: Use the fact that the countable product of two–point spaces with the product topology is homeomorphic to a Cantor set.

EXERCISE 1.9.5. Consider the *profinite* topology on $\mathbb{Z}$ in which open sets are defined as unions (not necessarily finite) of (non-constant and infinite in both directions) arithmetic progressions. Show that it is Hausdorff but not discrete.

EXERCISE 1.9.6. Let $\mathbb{T}^\infty$ be the product of countably many copies of the circle with the product topology. Define the map $\varphi : \mathbb{Z} \to \mathbb{T}^\infty$ by

$$\varphi(n) = (\exp(2\pi in/2), \exp(2\pi in/3), \exp(2\pi in/4), \exp(2\pi in/5), \dots)$$

Show that the map φ is injective and that the pullback topology on $\varphi(\mathbb{Z})$ coincides with its profinite topology.

EXERCISE 1.9.7. Prove that $\mathbb{R}$ (the real line) and $\mathbb{R}^2$ (the plane with the standard topology) are not homeomorphic.

Hint: Use the notion of connected set.

EXERCISE 1.9.8. Prove that the interior of any convex polygon in $\mathbb{R}^2$ is homeomorphic to $\mathbb{R}^2$.

EXERCISE 1.9.9. A topological space $(X, \mathcal{T})$ is called *regular* (or $(T3)$- space) if for any closed set $F \subset X$ and any point $x \in X \setminus F$ there exist disjoint open sets U and V such that $F \subset U$ and $x \in V$. Give an example of a Hausdorff topological space which is not regular.

EXERCISE 1.9.10. Give an example of a regular topological space which is not normal.

EXERCISE 1.9.11. Prove that any open convex subset of $\mathbb{R}^2$ is homeomorphic to $\mathbb{R}^2$.

EXERCISE 1.9.12. Prove that any compact topological space is sequentially compact.

EXERCISE 1.9.13. Prove that any sequentially compact topological space with countable base is compact.

EXERCISE 1.9.14. A point x in a topological space is called *isolated* if the one-point set $\{x\}$ is open. Prove that any compact separable Hausdorff space without isolated points contains a closed subset homeomorphic to the Cantor set.

EXERCISE 1.9.15. Find all different topologies (up to homeomorphism) on a set consisting of 4 elements which make it a connected topological space.

EXERCISE 1.9.16. Prove that the intersection of a nested sequence of compact connected subsets of a topological space is connected.

EXERCISE 1.9.17. Give an example of the intersection of a nested sequence of compact path connected subsets of a Hausdorff topological space which is not path connected.

EXERCISE 1.9.18. Let $A \subset \mathbb{R}^2$ be the set of all vectors (x, y) such that $x + y$ is a rational number and $x - y$ is an irrational number. Show that $\mathbb{R}^2 \setminus A$ is path connected.

EXERCISE 1.9.19. Prove that any compact one–dimensional manifold is homeomorphic to the circle.

EXERCISE 1.9.20. Prove that the Klein bottle is a compact topological manifold.

EXERCISE 1.9.21. Consider the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and let S be the quotient space obtained by identifying orbits of the map $I : x \mapsto -x$. Prove that S is homeomorphic to the sphere $\mathbb{S}^2$.

EXERCISE 1.9.22. Consider regular $2n$-gon and identify pairs of opposite side by the corresponding parallel translations. Prove that the identification space is a topological manifold.

EXERCISE 1.9.23. Prove that the manifolds obtained by this construction from the $4n$-gon and and $4n + 2$-gon are homeomorphic.

EXERCISE 1.9.24. Prove that the manifold of the previous exercise is homeomorphic to the surface of the sphere to which n "handles" are attached, or, equivalently, to the surface of n tori joint into a "chain" (Figure 1.8.1 illustrates this for $n = 1$ and $n = 3$.

EXERCISE 1.9.25. Let $f : \mathbb{S}^1 \to \mathbb{R}^2$ be a continuous map for which there are two points $a, b \in \mathbb{S}^1$ such that $f(a) = f(b)$ and f is injective on $\mathbb{S}^1 \setminus \{a\}$. Prove that $\mathbb{R}^2 \setminus f(\mathbb{S}^1)$ has exactly three connected components.

EXERCISE 1.9.26. Consider the one–parameter group of homeomorphisms of the real line generated by the map $x \to 2x$. Consider three separation properties: (T2), (T1), and

(T0) For any two points there exists an open set which contains one of them but not the other (but which one is not given in advance).

Which of these properties does the quotient topology possess?

EXERCISE 1.9.27. Consider the group $SL(2, \mathbb{R})$ of all 2×2 matrices with determinant one with the topolology induced from the natural coordinate embedding into $\mathbb{R}^4$. Prove that it is a topological group.

CHAPTER 2

METRICS AND RELATED STRUCTURES

The general notion of topology does not allow to compare neighborhoods of different points. Such a comparison is quite natural in various geometric contexts. The general setting for such a comparison is that of a *uniform structure*. The most common and natural way for a uniform structure to appear is via a metric, which was already mentioned on several occasions in Chapter 1, so we will postpone discussing the general notion of union structure to **??** until after detailed exposition of metric spaces. Another important example of uniform structures is that of topological groups, see Section 2.11 below in this chapter. Also, as in turns out, a Hausdorff compact space carries a natural uniform structure, which in the separable case can be recovered from any metric generating the topology. Metric spaces and topological groups are the notions central for foundations of analysis.

2.1. Definition of metric spaces and basic constructions

2.1.1. Axioms of metric spaces. We begin with listing the standard axioms of metric spaces, probably familiar to the reader from elementary real analysis courses, and mentioned in passing in Section 1.1, and then present some related definitions and derive some basic properties.

DEFINITION 2.1.1. If X is a set, then a function $d\colon X \times X \to \mathbb{R}$ is called a *metric* if

(1) $d(x, y) = d(y, x)$ *(symmetry)*,
(2) $d(x, y) \geq 0$; $d(x, y) = 0 \Leftrightarrow x = y$ *(positivity)*,
(3) $d(x, y) + d(y, z) \geq d(x, z)$ *(the triangle inequality)*.

If d is a metric, then (X, d) is called a *metric space*.

The set
$$B(x, r) := \{y \in X \mid d(x, y) < r\}$$
is called the *(open) r-ball* centered at x. The set
$$B_c(x, r) = \{y \in X \mid d(x, y) \leq r\}$$
is called the *closed r-ball* at (or around) x.

The *diameter* of a set in a metric space is the supremum of distances between its points; it is often denoted by diam A. The set A is called *bounded* if it has finite diameter.

A map $f : X \to Y$ between metric spaces with metrics d_X and d_Y is called as *isometric embedding* if for any pair of points $x, x' \in X$ $d_X(x, x') = d_Y(f(x), f(x'))$. If an isometric embedding is a bijection it is called an *isometry*. If there is an isometry between two metric spaces they are called *isometric*. This is an obvious equivalence relation in the category of metric spaces similar to homeomorphism for topological spaces or isomorphism for groups.

2.1.2. Metric topology. $O \subset X$ is called *open* if for every $x \in O$ there exists $r > 0$ such that $B(x, r) \subset O$. It follows immediately from the definition that open sets satisfy Definition 1.1.1. Topology thus defined is sometimes called the *metric topology* or *topology, generated by the metric* d. Naturally, different metrics may define the same topology. Often such metrics are called *equivalent*.

Metric topology automatically has some good properties with respect to bases and separation.

Notice that the closed ball $B_c(x, r)$ contains the closure of the open ball $B(x, r)$ but may not coincide with it (Just consider the integers with the the standard metric: $d(m, n) = |m - n|$.)

Open balls as well as balls or rational radius or balls of radius $r_n,\ n = 1, 2, \dots,$ where r_n converges to zero, form a base of the metric topology.

PROPOSITION 2.1.2. *Every metric space is first countable. Every separable metric space has countable base.*

PROOF. Balls of rational radius around a point form a base of neighborhoods of that point.

By the triangle inequality, every open ball contains an open ball around a point of a dense set. Thus for a separable spaces balls of rational radius around points of a countable dense set form a base of the metric topology. □

Thus, for metric spaces the converse to Proposition 1.1.12 is also true.

Thus the closure of $A \subset X$ has the form

$$\bar{A} = \{x \in X \mid \forall r > 0, \quad B(x, r) \cap A \neq \varnothing\}.$$

For any closed set A and any point $x \in X$ the *distance from* x *to* A,

$$d(x, A) := \inf_{y \in A} d(x, y)$$

is defined. It is positive if and only if $x \in X \setminus A$.

THEOREM 2.1.3. *Any metric space is normal as a topological space.*

PROOF. For two disjoint closed sets $A, B \in X$, let

$$\mathcal{O}_A := \{x \in X \mid d(x, A) < d(x, B),\ \mathcal{O}_B := \{x \in X \mid d(x, B) < d(x, A).$$

These sets are open, disjoint, and contain A and B respectively. □

Let $\varphi : [0, \infty] \to \mathbb{R}$ be a nondecreasing, continuous, concave function such that $\varphi^{-1}(\{0\}) = \{0\}$. If (X, d) is a metric space, then $\phi \circ d$ is another metric on d which generates the same topology.

It is interesting to notice what happens if a function d as in Definition 2.1.1 does not satisfy symmetry or positivity. In the former case it can be symmetrized producing a metric $d_S(x, y) := \max(d(x, y), d(y, x))$. In the latter by the symmetry and triangle inequality the condition $d(x,y) = 0$ defines an equivalence relation and a genuine metric is defined in the space of equivalence classes. Note that some of the most impotrant notions in analysis such as spaces L^p of functions on a measure space are actually not spaces of actual functions but are such quotient spaces: their elements are equivalence classes of functions which coincide outside of a set of measure zero.

2.1.3. Constructions.

1. *Inducing.* Any subset A of a metric space X is a metric space with an *induced metric* d_A, the restriction of d to $A \times A$.

2. *Finite products.* For the product of finitely many metric spaces, there are various natural ways to introduce a metric. Let $\varphi : ([0,\infty])^n \to \mathbb{R}$ be a continuous concave function such that $\varphi^{-1}(\{0\}) = \{(0,\dots,0)\}$ and which is nondecreasing in each variable.

Given metric spaces $(X_i, d_i),\ i = 1,\dots,n$, let

$$d^\varphi := \varphi(d_1,\dots,d_n) : (X_1 \times \dots X_n) \times (X_1 \times \dots X_n) \to \mathbb{R}.$$

EXERCISE 2.1.1. Prove that d^φ defines a metric on $X_1 \times \dots X_n$ which generates the product topology.

Here are examples which appear most often:

• the *maximum metric* corresponds to

$$\varphi(t_1,\dots,t_n) = \max(t_1,\dots,t_n);$$

• the l^p *metric* for $1 \le p < \infty$ corresponds to

$$\varphi(t_1,\dots,t_n) = (t_1^p + \dots + t_n^p)^{1/p}.$$

Two particularly important cases of the latter are $t = 1$ and $t = 2$; the latter produces the Euclidean metric in $\mathbb{R}^n$ from the standard (absolute value) metrics on n copies of $\mathbb{R}$.

3. *Countable products.* For a countable product of metric spaces, various metrics generating the product topology can also be introduced. One class of such metrics can be produced as follows. Let $\varphi : [0,\infty] \to \mathbb{R}$ be as above and let $a_1, a_2, \dots$ be a suquence of positive numbers such that the series $\sum_{n=1}^{\infty} a_n$ converges. Given metric spaces $(X_1, d_1), (X_2, d_2) \dots$, consider the metric d on the infinite product of the spaces $\{X_i\}$ defined as

$$d((x_1, x_2, \dots), (y_1, y_2, \dots)) := \sum_{n=1}^{\infty} a_n \varphi(d_n(x_n, y_n)).$$

EXERCISE 2.1.2. Prove that d is really a metric and that the corresponding metric topology coincides with the product topology.

4. *Factors.* On the other hand, projecting a metric even to a very good factor space is problematic. Let us begin with an example which exhibits some of the characteristic difficulties.

EXAMPLE 2.1.4. Consider the partition of the plane $\mathbb{R}^2$ into the level sets of the function xy, i.e. the hyperboli $xy = const \neq 0$ and the union of coordinate axes. The factor topology is nice and normal. It is easy to see in fact that the function xy on the factor space establishes a homeomorphism between this space and the real line. On the other hand, there is no natural way to define a metric in the factor space based on the Euclidean metric in the plane. Any two elements of the factor contain points arbitrary close to each other and arbitrary far away from each other so manipulating with infimums and supremums of of distances between the points in equivalence classes does not look hopeful.

We will see later that when the ambient space is compact and the factor-topology is Hausdorff there is a reasonable way to define a metric as the *Hausdorff metric* (see Definition 2.10.1) between equivalence classes considered as closed subsets of the space.

Here is a very simple but beautiful illustration how this may work.

EXAMPLE 2.1.5. Consider the real projective space $\mathbb{R}P(n)$ as the factor space of the sphere $\mathbb{S}^n$ with opposite points identified. Define the distance between the pairs $(x, -x)$ and $(y, -y)$ as the minimum of distances between members of the pairs. Notice that this minimum is achieved simultaneously on a pair and the pair of opposite points. This last fact allows to check the triangle inequality (positivity and symmetry are obvious) which in general would not be satisfied for the minimal distance of elements of equivalence classes even if those classes are finite.

EXERCISE 2.1.3. Prove the triangle inequality for this example. Prove that the natural projection from $\mathbb{S}^n$ to $\mathbb{R}P(n)$ is an isometric embedding in a neighborhood of each point. Calculate the maximal size of such a neighborhood.

Our next example is meant to demonstrate that the chief reason for the success of the previous example is not compactness but the fact that the factor space is the orbit space of an action by isometries (and of course is Hausdorff at the same time):

EXAMPLE 2.1.6. Consider the natural projection $\mathbb{R}^n \to \mathbb{R}^n/\mathbb{Z}^n = \mathbb{T}^n$. Define the distance $d(a\mathbb{Z}^n, b\mathbb{Z}^n)$ on the torus as the minimum of Euclidean distances between points in $\mathbb{R}^n$ in the equivalence classes representing corresponding points on the torus. Notice that since translations are isometries the minimum is always achieved and if it is achieved on a pair (x, y) it is also achieved on any integer translation of (x, y).

EXERCISE 2.1.4. Prove the triangle inequality for this example. Prove that the natural projection from $\mathbb{R}^n$ to $\mathbb{T}^n$ is an isometric embedding in any open ball of radius 1/2 and is not an isometric embedding in any open ball of any greater radius.

2.2. Cauchy sequences and completeness

2.2.1. Definition and basic properties. The notion of Cauchy sequence in Euclidean spaces and the role of its convergence should be familiar from elementary real analysis courses. Here we will review this notion in the most general setting, leading up to general theorems on completion, which play a crucial role in functional analysis.

DEFINITION 2.2.1. A sequence $\{x_i\}_{i\in\mathbb{N}}$ is called a *Cauchy sequence* if for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d(x_i, x_j) < \varepsilon$ whenever $i, j \geq \mathbb{N}$; X is said to be *complete* if every Cauchy sequence converges.

PROPOSITION 2.2.2. *A subset A of a complete metric space X is a complete metric space with respect to the induced metric if and only if it is closed.*

PROOF. For a closed $A \in X$ the limit of any Cauchy sequence in A belongs to A. If A is not closed take a sequence in A converging to a point in $\bar{A} \setminus A$. It is Cauchy but does not converge in A. □

The following basic property of complete spaces is used in the next two theorems.

PROPOSITION 2.2.3. *Let $A_1 \supset A_2 \supset \dots$ be a nested sequence of closed sets in a complete metric space, such that* $\operatorname{diam} A_n \to 0$ *as $n \to \infty$. Then $\bigcap_{n=1}^{\infty} A_n$ is a single point.*

PROOF. Since $\operatorname{diam} A_n \to 0$ the intersection cannot contain more than one point. Take a sequence $x_n \in A_n$. It is Cauchy since $\operatorname{diam} A_n \to 0$. Its limit x belongs to $\bar{A}_n$ for any n. Since the sets A_i are closed, it follows that $x \in A_n$ for any n. □

2.2.2. The Baire category theorem.

THEOREM 2.2.4 (Baire Category Theorem). *In a complete metric space, a countable intersection of open dense sets is dense. The same holds for a locally compact Hausdorff space.*

PROOF. If $\{O_i\}_{i\in\mathbb{N}}$ are open and dense in X and $\varnothing \neq B_0 \subset X$ is open then inductively choose a ball B_{i+1} of radius at most ε/i for which we have $\bar{B}_{i+1} \subset O_{i+1} \cap B_i$. The centers converge by completeness, so

$$\varnothing \neq \bigcap_i \bar{B}_i \subset B_0 \cap \bigcap_i O_i.$$

For locally compact Hausdorff spaces take B_i open with compact closure and use the finite intersection property. □

The Baire Theorem motivates the following definition. If we want to mesure massivenes of sets in a topological or in particular metric space, we may assume that nowhere dense sets are small and their complements are massive. The next natural step is to introduce the following concept.

DEFINITION 2.2.5. Countable unions of nowhere dense sets are called *sets of first (Baire) category*.

The complement to a set of first baire category is called a *residual* set.

The Baire category theorem asserts that, at least for complete metric spaces, sets of first category can still be viewed as small, since they cannot fill any open set.

The Baire category theorem is a simple but powerful tool for proving *existence* of various objects when it is often difficult or impossible to produce those constructively.

2.2.3. Minimality of the Cantor set. Armed with the tools developed in the previous subsections, we can now return to the Cantor set and prove a universality theorem about this remarkable object.

THEOREM 2.2.6. *(cf. Exercise 1.9.14)*
Any uncountable separable complete metric space X contains a closed subset homeomorphic to the Cantor set.

PROOF. First consider the following subset

$$X_0 \colon \{x \in X | \text{any neigbourhood of } x \text{ contains uncountably many points}\}$$

Notice that the set X_0 is perfect, i.e., it is closed and contains no isolated points.

LEMMA 2.2.7. *The set $X \setminus X_0$ is countable.*

PROOF. To prove the lemma, for each point $x \in X \setminus X_0$ find a neighborhood from a countable base which contains at most countably many points (Proposition 2.1.2). Thus $X \setminus X_0$ is covered by at most countably many sets each containing at most countably many points. □

Thus the theorem is a consequence of the following fact.

PROPOSITION 2.2.8. *Any perfect complete metric space X contains a closed subset homeomorphic to the Cantor set.*

PROOF. To prove the the proposition, pick two points $x_0 \neq x_1$ in X and let $d_0 := d(x_0, x_1)$. Let

$$X_i := \overline{B(x_i, (1/4)d_0)}, \quad i = 0, 1$$

and $C_1 := X_0 \cup X_1$.

Then pick two different points $x_{i,0}, x_{i,1} \in \operatorname{Int} X_i,\ i = 0, 1$. Such choices are possible because any open set in X contains infinitely many points. Notice that $d(x_{i,0}, x_{i,1}) \le (1/2)d_0$. Let

$$Y_{i_1,i_2} := \overline{B(x_{i_1,i_2}, (1/4)d(x_{i_1,0}, x_{i_1,1}))},\ i_1, i_2 = 0, 1,$$

$$X_{i_1,i_2} := Y_{i_1,i_2} \cap C_1 \quad \text{and } C_2 = X_{0,0} \cup X_{0,1} \cup X_{1,0} \cup X_{1,1}.$$

Notice that $\operatorname{diam}(X_{i_1,i_2}) \le d_0/2$.

Proceed by induction. Having constructed

$$C_n = \bigcup_{i_1,\dots,i_n \in \{0,1\}} X_{i_1,\dots,i_n}$$

with $\operatorname{diam} X_{i_1,\dots,i_n} \le d_0/2^n$, pick two different points $x_{i_1,\dots,i_n,0}$ and $x_{i_1,\dots,i_n,1}$ in $\operatorname{Int} X_{i_1,\dots,i_n}$ and let us successively define

$$Y_{i_1,\dots,i_n,i_{n+1}} := \bar{B}(x_{i_1,\dots,i_n,i_{n+1}}, d(x_{i_1,\dots,i_n,0}, x_{i_1,\dots,i_n,1})/4),$$

$$X_{i_1,\dots,i_n,i_{n+1}} := Y_{i_1,\dots,i_n,i_{n+1}} \cap C_n,$$

$$C_{n+1} := \bigcup_{i_1,\dots,i_n,i_{n+1} \in \{0,1\}} X_{i_1,\dots,i_n,i_{n+1}}.$$

Since $\operatorname{diam} X_{i_1,\dots,i_n} \le d_0/2^n$, each infinite intersection

$$\bigcap_{i_1,\dots,i_n,\dots \in \{0,1\}} X_{i_1,\dots,i_n,\dots}$$

is a single point by Heine–Borel (Proposition 2.2.3). The set $C := \bigcap_{n=1}^{\infty} C_n$ is homeomorphic to the countable product of the two point sets $\{0, 1\}$ via the map

$$\bigcap_{i_1,\dots,i_n,\dots \in \{0,1\}} X_{i_1,\dots,i_n,\dots} \mapsto (i_1, \dots, i_n \dots).$$

By Proposition 1.7.3, C is homeomorphic to the Cantor set. □

The theorem is thus proved. □

2.2.4. Completion. Completeness allows to perform limit operations which arise frequently in various constructions. Notice that it is not possible to define the notion of Cauchy sequence in an arbitrary topological space, since one lacks the possibility of comparing neighborhoods at different points. Here the uniform structure (see **??**) provides the most general natural setting.

A metric space can be made complete in the following way:

DEFINITION 2.2.9. If X is a metric space and there is an isometry from X onto a dense subset of a complete metric space $\hat{X}$ then $\hat{X}$ is called the *completion* of X.

THEOREM 2.2.10. *For any metric space X there exists a completion unique up to isometry which commutes with the embeddings of X into a completion.*

PROOF. The process mimics the construction of the real numbers as the completion of rationals, well–known from basic real analysis. Namely, the elements of the completion are equivalence classes of Cauchy sequences by identifying two sequences if the distance between the corresponding elements converges to zero. The distance between two (equivalence classes of) sequences is defined as the limit of the distances between the corresponding elements. An isometric embedding of X into the completion is given by identifying element of X with constant sequences. Uniqueness is obvious by definition, since by uniform continuity the isometric embedding of X to any completion extends to an isometric bijection of the standard completion. □

2.3. The p-adic completion of integers and rationals

This is an example which rivals the construction of real numbers in its importance for various areas of mathematics, especially to number theory and algebraic geometry. Unlike the construction of the reals, it gives infinitely many differnt nonisometric completions of the rationals.

2.3.1. The p-adic norm. Let p be a positive prime number. Any rational number r can be represented as $p^m \frac{k}{l}$ where m is an integer and k and l are integers realtively prime with p. Define the *p-adic norm* $\|r\|_p := p^{-m}$ and the distance $d_p(r_1, r_2) := \|r_1 - r_2\|_p$.

EXERCISE 2.3.1. Show that the p-adic norm is *multiplicative*, i.e., we have $\|r_1 \cdot r_2\|_p = \|r_1\|_p \|r_2\|_p$.

PROPOSITION 2.3.1. *The inequality*

$$d_p(r_1, r_3) \leq \max(d_p(r_1, r_2), d_p(r_2, r_3))$$

holds for all $r_1, r_2, r_3 \in \mathbb{Q}$.

REMARK 2.3.2. A metric satisfying this property (which is stronger than the triangle inequality) is called an *ultrametric*.

PROOF. Since $\|r\|_p$= $\| - r\|_p$ the statement follows from the property of p-norms:

$$\|r_1 + r_2\|_p \leq \|r_1\|_p + \|r_2\|_p.$$

To see this, write $r_i = p_i^m \frac{k_i}{l_i}$, $i = 1, 2$ with k_i and l_i relatively prime with p and assume without loss of generality that $m_2 \geq m_1$. We have

$$r_1 + r_2 = p_1^m \frac{k_1 l_2 + p^{m_2 - m_1} k_2 l_1}{l_1 l_2}.$$

The numerator $k_1 l_2 + p^{m_2 - m_1} k_2 l_1$ is an integer and if $m_2 > m_1$ it is relatively prime with p. In any event we have $\|r_1 + r_2\|_p \leq p^{-m_1} = \|r_1\|_p = \max(\|r_1\|_p, \|r_2\|_p)$. □

2.3.2. The p-adic numbers and the Cantor set. Proposition 2.3.1 and the multiplicativity prorerty of the p-adic norm allow to extend addition and multiplication from $\mathbb{Q}$ to the completion. This is done in exacly the same way as in the real analysis for real numbers. The existence of the opposite and inverse (the latter for a nonzero element) follow easily.

Thus the completion becomes a field, which is called the *field of p-adic numbers* and is usually denoted by $\mathbb{Q}_p$. Restricting the procedure to the integers which always have norm ≤ 1 one obtains the subring of $\mathbb{Q}_p$, which is called the *ring of p-adic integers* and is usually denoted by $\mathbb{Z}_p$.

The topology of p–adic numbers once again indicates the importance of the Cantor set.

PROPOSITION 2.3.3. *The space $\mathbb{Z}_p$ is homeomorphic to the Cantor set; $\mathbb{Z}_p$ is the unit ball (both closed and open) in $\mathbb{Q}_p$.*

The space $\mathbb{Q}_p$ is homeomorphic to the disjoint countable union of Cantor sets.

PROOF. We begin with the integers. For any sequence

$$a = \{a_n\} \in \prod_{n=1}^{\infty} \{0, 1 \ldots, p-1\}$$

the sequence of integers

$$k_n(a) := \sum_{i=1}^{n} a_n p^i$$

is Cauchy; for different $\{a_n\}$ these sequences are non equivalent and any Cauchy sequence is equivalent to one of these. Thus the correspondence

$$\prod_{n=1}^{\infty} \{0, 1 \ldots, p-1\} \to \mathbb{Z}_p, \quad \{a_n\} \mapsto \text{the equivalence class of } k_n(a)$$

is a homeomorphism. The space $\prod_{n=1}^{\infty} \{0, 1 \ldots, p-1\}$ can be mapped homeomorphically to a nowhere dense perfect subset of the interval by the map

$$\{a_n\}_{n=1}^{\infty} \mapsto \sum_{n=1}^{\infty} a_n (2p-1)^{-i}$$

. Thus the statement about $\mathbb{Z}_p$ follows from Proposition 1.7.5.

Since $\mathbb{Z}$ is the unit ball (open and closed) around 0 in the matric d_p and any other point is at a distance at least 1 from it, the same holds for the completions.

Finally, any rational number can be uniquely represented as

$$k + \sum_{i=1}^{n} a_i p^{-i}, \quad k \in \mathbb{Z}, \quad a_i \in \{0, \ldots, p-1\}, \ i = 1, \ldots, n.$$

If the corresponging finite sequences a_i have different length or do not coincide, then the p-adic distance between the rationals is at least 1. Passing to the completion we see that any $x \in \mathbb{Q}_p$ is uniquely represented as $k + \sum_{i=1}^{n} a_i p^{-i}$ with $k \in \mathbb{Z}_p$. with pairwise distances for different a_i's at least one. $\square$

EXERCISE 2.3.2. Where in the construction is it important that p is a prime number?

2.4. Maps between metric spaces

2.4.1. Stronger continuity properties.

DEFINITION 2.4.1. A map $f : X \to Y$ between the metric spaces (X, d), (Y, dist) is said to be *uniformly continuous* if for all $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x, y \in X$ with $d(x, y) < \delta$ we have $\mathrm{dist}(f(x), f(y)) < \varepsilon$. A uniformly continuous bijection with uniformly continuous inverse is called a *uniform homeomorphism*.

PROPOSITION 2.4.2. *A uniformly continuous map from a subset of a metric space to a complete space uniquely extends to its closure.*

PROOF. Let $A \subset X$, $x \in \bar{A}$, $f\colon A \to Y$ uniformly continuous. Fix an $\epsilon > 0$ and find the corresponding δ from the definition of uniform continuity. Take the closed $\delta/4$ ball around x. Its image and hence the closure of the image has diameter $\leq \epsilon$. Repeating this procedure for a sequence $\epsilon_n \to 0$ we obtain a nested sequence of closed sets whose diameters converge to zero. By Proposition 2.2.3 their intersection is a single point. If we denote this point by $f(x)$ the resulting map will be continuous at x and this extension is unique by uniqueness of the limit since by construction for any sequence $x_n \in A,\ x_n \to x$ one has $f(x_n) \to f(x)$. □

DEFINITION 2.4.3. A family $\mathcal{F}$ of maps $X \to Y$ is said to be *equicontinuous* if for every $x \in X$ and $\varepsilon > 0$ there is a $\delta > 0$ such that $d(x, y) < \delta$ implies

$$\mathrm{dist}(f(x), f(y)) < \varepsilon \quad \text{for all } y \in X \text{ and } f \in \mathcal{F}.$$

DEFINITION 2.4.4. A map $f\colon X \to Y$ is said to be *Hölder continuous* with exponent α, or α-*Hölder*, if there exist $C, \varepsilon > 0$ such that $d(x, y) < \varepsilon$ implies

$$d(f(x), f(y)) \leq C(d(x, y))^\alpha,$$

Lipschitz continuous if it is 1-Hölder, and biLipschitz if it is Lipschitz and has a Lipschitz inverse.

For a Lipschitz map f infimum of all C for which the inequality $d(f(x), f(y)) \leq C(d(x, y))$ holds is called the *Lipschitz constant* of f.

It is useful to introduce local versions of the above notions. A map $f\colon X \to Y$ is said to be Hölder continuous with exponent α, at the point $x \in X$ or α-*Hölder*, if there exist $C, \varepsilon > 0$ such that $d(x, y) < \varepsilon$ implies

$$d(f(x), f(y)) \leq C(d(x, y))^\alpha,$$

Lipschitz continuous at x if it is 1-Hölder at x.

2.4.2. Various equivalences of metric spaces. Besides the natural relation of isometry, the category of metric spaces is endowed with several other equivalence relations.

DEFINITION 2.4.5. Two metric spaces are *uniformly equivalent* if there exists a homeomorphism between the spaces which is uniformly continuous together with its inverse.

PROPOSITION 2.4.6. *Any metric space uniformly equivalent to a complete space is complete.*

PROOF. A uniformly continuous map obviously takes Cauchy sequences to Cauchy sequences. □

EXAMPLE 2.4.7. The open interval and the real line are homeomorphic but not uniformly equivalent because one is bounded and the other is not.

EXERCISE 2.4.1. Prove that an open half–line is not not uniformly equivalent to either whole line or an open interval.

DEFINITION 2.4.8. Metric spaces are *Hölder equivalent* if there there exists a homeomorphism between the spaces which is Hölder together with its inverse.

Metric spaces are *Lipschitz equivalent* if there exists a biLipschitz homeomorphism between the spaces.

EXAMPLE 2.4.9. Consider the standard middle–third Cantor set C and the subset C_1 of $[0, 1]$ obtained by a similar procedure but with taking away at every step the open interval in the middle of one half of the length. These two sets are Hólder equivalent but not Lipschitz equivalent.

EXERCISE 2.4.2. Find a Hölder homeomorphism with Hölder inverse in the previous example.

As usual, it is easier to prove existence of an equivalence that absence of one. For the latter one needs to produce an invariant of Lipschitz equivalence calculate it for two sets and show that the values (which do not have to be numbers but may be mathematical objects of another kind) are different. On this occasion one can use asymptotics of the minimal number of ϵ-balls needed to cover the set as $\epsilon \to 0$. Such notions are called *capacities* and are related to the important notion of *Hausdorff dimension* which, unlike the topological dimension, is not invariant under homeomorphisms. See **??**.

EXERCISE 2.4.3. Prove that the identity map of the product space is biLIpschitz homeomorphism between the space provided with the maximal metric and with any l^p metric.

EXAMPLE 2.4.10. The unit square (open or closed) is Lipschitz equivalent to the unit disc (respectively open or closed), but not isometric to it.

EXERCISE 2.4.4. Consider the unit circle with the metric induced from the $\mathbb{R}^2$ and the unit circle with the angular metric. Prove that these two metric spaces are Lipschitz equivalent but not isometric.

2.4.3. Contraction mapping principle.

DEFINITION 2.4.11. Let (X, d) be a metric space. A map $f\colon X \to X$ is said to be *contracting* if there exists $\lambda < 1$ such that for any $x, y \in X$

$$d(f(x), f(y)) \le \lambda d(x, y). \tag{2.4.1}$$

Notice that the infimum of numbers λ satisfying (2.4.1) also satisfies this condition. This justifies calling this number the *contraction coefficient* of f. It is in fact the Lipschitz constant (Definition 2.4.4) of f. It is positive unless f maps the whole space into a single point. Thus one can say that a map is contracting if it Lipschitz continuous with Lipschitz constant less than one.

DEFINITION 2.4.12. We say that two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ of points in a metric space *converge exponentially* (or *with exponential speed*) *to each other* if $d(x_n, y_n) < c\lambda^n$ for some $c > 0$, $\lambda < 1$. In particular, if one of the sequences is constant, that is, $y_n = y$, we say that x_n *converges exponentially* to y.

PROPOSITION 2.4.13 (Contraction Mapping Principle). *Let X be a complete metric space. Under the action of iterates of a contracting map $f\colon X \to X$ all points converge with exponential speed to the unique fixed point of f.*

Thus for a contracting map all points are asymptotic to a unique fixed point.

PROOF. Iteration gives

$$d(f^n(x), f^n(y)) \le \lambda^n d(x, y)$$

for $n \in \mathbb{N}$, so

$$d(f^n(x), f^n(y)) \to 0 \quad \text{as } n \to \infty.$$

This means that the asymptotic behavior of all points is the same. On the other hand, for any $x \in X$ the sequence $\{f^n(x)\}_{n\in\mathbb{N}}$ is a Cauchy sequence because if $m \ge n$ then

$$\begin{aligned} d(f^m(x), f^n(x)) &\le \sum_{k=0}^{m-n-1} d(f^{n+k+1}(x), f^{n+k}(x)) \\ &\le \sum_{k=0}^{m-n-1} \lambda^{n+k} d(f(x), x) \le \frac{\lambda^n}{1-\lambda} d(f(x), x) \xrightarrow[n\to\infty]{} 0. \end{aligned} \tag{2.4.2}$$

Thus, $p := \lim_{n\to\infty} f^n(x)$ exists if the space is complete. By (2.4.1) this limit is the same for all x and p is a fixed point because

$$p = \lim_{n\to\infty} f^n(x) = \lim_{n\to\infty} f^{n+1}(x) = \lim_{n\to\infty} f(f^n(x)) = f(\lim_{n\to\infty} f^n(x)) = f(p).$$

By (2.4.1) there is at most one fixed point. Letting $m \to \infty$ in (2.4.2) gives $d(f^n(x), p) \leq \dfrac{\lambda^n}{1-\lambda} d(f(x), x)$. □

COROLLARY 2.4.14. *Let f be a continuous map of a (not necessarily complete) metric space such that for an equivalent complete metric f is a contraction. Than f has unique fixed point.*

Contraction Mapping Principle which we have just proved is, despite the great simplicity of its proof, one of the most useful and most widely used tools of mathematical analysis. It is used in the standard proofs of such basic but fundamental facts as the Implicit Function theorem, existence and uniqueness of solutions of ODE, to more advance but still central results as the Stable Manifold Theorem, The Center Manifold Theorem, to sophisticated existence theorems in PDE and differential geometry. Usually the main work goes into constructing an appropriate space, proving its completeness and obtaining estimates which guarantee contraction property, not necessarily with respect to the original metric but with respect to some equivalent metric. Now we will illustrate usefulness of Contraction Mapping Principle by one simple but important application.

PROPOSITION 2.4.15. *If p is a periodic point of period m for a C^1 map f and the differential Df^m_p does not have one as an eigenvalue then for every map g sufficiently close to f in the C^1 topology there exists a unique periodic point of period m close to p.*

PROOF. We introduce local coordinates near p with p as the origin. In these coordinates Df^m_0 becomes a matrix. Since 1 is not among its eigenvalues the map $F = f^m - \mathrm{Id}$ defined locally in these coordinates is locally invertible by the Inverse Function Theorem. Now let g be a map C^1-close to f. Near 0 one can write $g^m = f^m - H$ where H is small together with its first derivatives. A fixed point for g^m can be found from the equation $x = g^m(x) = (f^m - H)(x) = (F + \mathrm{Id} - H)(x)$ or $(F - H)(x) = 0$ or

$$x = F^{-1}H(x).$$

Since F^{-1} has bounded derivatives and H has small first derivatives one can show that $F^{-1}H$ is a contracting map. More precisely, let $\| \cdot \|_0$ denote the C^0-norm, $\|dF^{-1}\|_0 = L$, and suppose $\max(\|H\|_0, \|dH\|_0) \leq \epsilon$. Then, since $F(0) = 0$, we get $\|F^{-1}H(x) - F^{-1}H(y)\| \leq \epsilon L \|x - y\|$ for every x, y close to 0 and $\|F^{-1}H(0)\| \leq L\|H(0)\| \leq \epsilon L$, so

$$\|F^{-1}H(x)\| \leq \|F^{-1}H(x) - F^{-1}H(0)\| + \|F^{-1}H(0)\| \leq \varepsilon L \|x\| + \varepsilon L.$$

Thus if $\epsilon \leq \dfrac{R}{L(1+R)}$ the disc $X := \{x \mid \|x\| \leq R\}$ is mapped by $F^{-1}H$ into itself and the map $F^{-1}H \colon X \to X$ is contracting. By the Contraction Mapping Principle it has a unique fixed point in X which is thus a unique fixed point for g^m near 0. □

2.5. Role of metrics in geometry and topology

2.5.1. Elementary geometry. The study of metric spaces with a given metric belongs to the realm of geometry. The natural equivalence relation here is the strongest one, mentioned above, the isometry. Recall that the classical (or "elementary") Euclidean geometry deals with properties of simple objects in the plane or in the three-dimensional space invariant under isometries, or, according to some interpretations, under a larger class of similarity transformations since the absolute unit of length is not fixed in the Euclidean geometry (unlike the prototype non-Euclidean geometry, the hyperbolic one!).

Isometries tend to be rather rigid: recall that in the Euclidean plane an isometry is uniquely determined by images of three points (not on a line) , and in the Euclidean space by the images of four (not in a plane), and those images cannot be arbitrary.

EXERCISE 2.5.1. Prove that an isometry of $\mathbb{R}^n$ with the standard Euclidean metric is uniquely determined by images of any points $x_1, \dots, x_{n+1}$ such that the vectors $x_k - x_1,\ k = 2, \dots, n+1$ are linearly independent.

2.5.2. Riemannian geometry. The most important and most central for mathematics and physics generalization of Euclidean geometry is *Riemannian geometry*. Its objects are manifolds (in fact, differentiable or smooth manifoldswith an extra structure of a *Riemannian metric* which defines Euclidean geometry (distances and angles) *infinitesimally* at each point, and the length of curves is obtained by integration. A smooth manifolds with a fixed Riemannian metric is called a *Riemannian manifold*. Instances of it have already appeared, e.g. the metric on the standard embedded sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ where the distance is measured along the great circles, (and is *not* induced from $\mathbb{R}^{n+1}$), its projection to $\mathbb{R}P(n)$, and projection of Euclidean metric in $\mathbb{R}^n$ to the torus $\mathbb{T}^n$.

EXERCISE 2.5.2. Prove that in the spherical geometry the sum of angels of a triangle whose sides are arcs of great circles is always greater than π

2.5.3. More general metric geometries. Riemannian geometry is the richest and the most important but by no means only and not the most general way metric spaces appear in geometry. While Riemannian geometry, at least classically, has been inspired mostly by analytic methods of classical geometries (Euclidean, spherical and suchlike) there are other more contemporary directions which to a large extent are developing the synthetic methods of classical geometric reasoning; an outstanding example is the geometry of *Aleksandrov spaces*.

EXERCISE 2.5.3. Let $a > 0$ and denote by C_a the surface of the cone in $\mathbb{R}^3$ given by the conditions $a^2z^2 = x^2 + y^2$, $z \geq 0$. Call a curve in C_a a line segment if it is the shortest curve between its endpoints. Find all line segments in C_a.

2.5.4. Metric as a background and a base for other structures. The most classical extensions of Euclidean geometry dealt (with the exception of spherical geometry) not with other metrics spaces but with geometric structures more general than Euclidean metric, such as affine and projective structures. To this one should add conformal structure which is of central importance for complex analysis. In all these geometries metrics appear in an auxiliary role such as the metric from Example 2.1.5 on real projective spaces.

EXERCISE 2.5.4. Prove that there is no metric on the projective line $\mathbb{R}P(1)$ generating the standard topology which is invariant under projective transformations.

EXERCISE 2.5.5. Prove that there is no metric in $\mathbb{R}^2$ generating the standard topology and invariant under all area preserving affine transformations, i.e transformations of the form $x \mapsto Ax + b$ where A is a matrix with determinanat ± 1 and b is a vector.

The role played by metrics in the principal branches of topology, algebraic and differential topology, is somewhat similar. Most spaces studied in those disciplines are metrizable; especially in the case of differential topology which studies smooth manifolds and various derivative objects, fixing a Riemannian metric on the manifold is very useful. It allows to bring precise measurements into the picture and provides various function spaces associated with the manifold such as spaces of smooth functions or differential forms, with the structure of a Banach space. But the choice of metric is usually arbitrary and only in the special cases, when the objects of study possess many symmetries, a particular choice of metric sheds much light on the core topological questions.

One should also point out that in the study of non-compact topological spaces and group actions on such spaces often a natural class of biLipschitz equivalent metrics appear. The study of such structures has gained importance over last two decades.

2.6. Separation properties and metrizability

As we have seen any metric topology is first countable (Proposition 2.1.2) and normal (Theorem 2.1.3). Conversely, it is natural to ask under what conditions a topological space has a metric space structure compatible with its topology.

A topological space is said to be *metrizable* if there exists a metric on it that induces the given topology. The following theorem gives necessary and sufficient conditions for metrizability for second countable topological spaces.

THEOREM 2.6.1. *[Urysohn Metrization Theorem]*
A normal space with a countable base for the topology is metrizable.

Theorem 2.6.1 and Proposition 1.5.4 imply

COROLLARY 2.6.2. *Any compact Hausdorff space with a countable base is metrizable.*

2.7. Compact metric spaces

2.7.1. Sequential compactness.

PROPOSITION 2.7.1. *Any compact metric space is complete.*

PROOF. Suppose the opposite, that is, X is a compact metric space and a Cauchy sequence $x_n,\ n = 1, 2, \dots$ does not converge. By taking a subsebuence if necessary we may assume that all points x_n are different. The union of the elements of the sequence is closed since the sequence does not converge. Let

$$\mathcal{O}_n := X \setminus \bigcup_{i=n}^{\infty} \{x_n\}.$$

These sets form an open cover of X but since they are increasing there is no finite subcover. □

DEFINITION 2.7.2. Given $r > 0$ a subset A of a metric space X is called an *r-net* if for any $x \in X$ there is $a \in A$ such that the distance $d(x, a)$. Equivalently r-balls around the points of A cover X.

A set $A \subset X$ is called *r-separated* if the distance between any two different points in A is greater than r.

The following observation is very useful in the especially for quantifying the notion of compactness.

PROPOSITION 2.7.3. *Any maximal r-separated set is an r-net.*

PROOF. If A is r-separated and is not an r-net then there is a point $x \in X$ at a distance $\geq r$ from every point of A Hence the set $A \cup \{x\}$ is r-separated □

PROPOSITION 2.7.4. *The following properties of a metric space X are equivalent*

(1) *X is compact;*
(2) *for any $\epsilon > 0$ X contains a finite ϵ-net, or, equivalently, any r-separated set for any $r > 0$ is finite;*
(3) *every sequence contains a congerving subsequence.*

PROOF. (1) → (2). If X is compact than the cover of X by all balls of radius ϵ contains a finite subcover; centers of those balls form a finite ϵ-net.

(2) → (3) By Proposition 2.7.1 it is sufficient to show that every sequence has a Cauchy subsequence. Take a sequence $x_n,\ n = 1, 2, \dots$ and consider a finite 1-net. There is a ball of radius 1 which contains infinitely many elements of the sequence. Consider only these elements as a subsequence. Take a finite 1/2-net and find a subsequence which lies in a single ball of radius 1/2. Continuing by induction we find nested subsequences of the original sequence which lie in balls of radius $1/2^n$. Using the standard diagonal process we construct a Cauchy subsequence.

(3) → (1). Let us first show that the space must be separable. This implies that any cover contains a countable subcover since the space has countable base. If the

space is not separable than there exists an $\epsilon > 0$ such that for any countable (and hence finite) collection of points there is a point at the distance greater than ϵ from all of them. This allows to construct by induction an infinite sequence of points which are pairwise more than ϵ apart. Such a sequence obviously does not contain a converging subsequence.

Now assume there is an open countable cover $\{\mathcal{O}_1, \mathcal{O}_2, \dots\}$ without a finite subcover. Take the union of the first n elements of the cover and a point x_n outside of the union. The sequence $x_n,\ n = 1, 2, \dots$ thus defined has a converging subsequence $x_{n_k} \to x$. But x belong to a certain element of the cover, say $\mathcal{O}_N$. Then for a sufficinetly large k, $n_k > N$ hence $x_{n_k} \notin \mathcal{O}_N$, a contradiction to convergence. □

An immediate corollary of the proof is the following.

PROPOSITION 2.7.5. *Any compact metric space is separable.*

Aside from establishing equivalence of compactness and sequential compactness for metric spaces Proposition 2.7.4 contains a very useful criterion of compactness in the form of property (2). Right away it gives a necessary and sufficient condition for a (in general incomplete) metric space to have compact completion. As we see it later in Section 2.7.5 it is also a starting point for developing qualitative notions related to the "size" of a metric space.

DEFINITION 2.7.6. A metric space (X, d) is *totally bounded* if it contains a finite ϵ-net for any $\epsilon > 0$, or, equivalently if any r-separates subset of X for any $r > 0$ is finite.

Since both completion and any subset of a totally bounded space are totally bounded Proposition 2.7.4 immediately implies

COROLLARY 2.7.7. *Completion of a metric space is compact if and only if the space is totally bounded.*

EXERCISE 2.7.1. Prove that an isometric embedding of a compact metric space into itself is an isometry.

2.7.2. Lebesgue number.

PROPOSITION 2.7.8. *For an open cover of a compact metric space there exists a number δ such that every δ-ball is contained in an element of the cover.*

PROOF. Suppose the opposite. Then there exists a cover and a sequence of points x_n such that the ball $B(x_n, 1/2^n)$ does not belong to any element of the cover. Take a converging subsequence $x_{n_k} \to x$. Since the point x is covered by an open set, a ball of radius $r > 0$ around x belongs to that element. But for k large enough $d(x, x_{n_k}) < r/2$ and hence by the triangle inequality the ball $B(x_{n_k}, r/2)$ lies in the same element of the cover. □

The largest such number is called the *Lebesgue number* of the cover.

2.7.3. Characterization of Cantor sets.

THEOREM 2.7.9. *Any perfect compact totally disconnected metric space X is homeomorphic to the Cantor set.*

PROOF. Any point $x \in X$ is contained in a set of arbitrally small diameter which is both closed and open. For x is the intersection of all sets which are open and closed and contain x. Take a cover of $X \setminus X$ by sets which are closed and open and do not contain x Adding the ball $B(x, \epsilon)$ one obtains a cover of X which has a finite subcover. Union of elements of this subcover other than $B(x, \epsilon)$ is a set which is still open and closed and whose complement is contained in $B(x, \epsilon)$.

Now consider a cover of the space by sets of diameter ≤ 1 which are closed and open. Take a finite subcover. Since any finite intersection of such sets is still both closed an open by taking all possible intersection we obtain a *partition* of the space into finitely many closed and open sets of diameter ≤ 1. Since the space is perfect no element of this partition is a point so a further division is possible. Repeating this procedure for each set in the cover by covering it by sets of diameter $\leq 1/2$ we obtain a finer partition into closed and open sets of of diameter $\leq 1/2$. Proceeding by induction we obtain a nested sequence of finite partitions into closed and open sets of positive diameter $\leq 1/2^n$, $n = 0, 1, 2, \ldots$. Proceeding as in the proof of Proposition 1.7.5, that is, mapping elements of each partition inside a nested sequence of contracting intervals, we constuct a homeomorphism of the space onto a nowhere dense perfect subset of $[0, 1]$ and hence by Proposition 1.7.5 our space is homeomorphic to the Cantor set. □

2.7.4. Universality of the Hilbert cube. Theorem 2.2.6 means that Cantor set is in some sense a minimal nontrivial compact metrizable space. Now we will find a maximal one.

THEOREM 2.7.10. *Any compact separable metric space X is homeomorphic to a closed subset of the Hilbert cube H.*

PROOF. First by multiplying the metric by a constant if nesessary we may assume that the diameter of X is less that 1. Pick a dense sequence of points $x_1, x_2 \ldots$ in X. Let $F : X \to H$ be defined by

$$F(x) = (d(x, x_1), d_(x, x_2), \ldots).$$

This map is injective since for any two distict points x and x' one can find n such that $d(x, x_n) < (1/2)d(x', x_n)$ so that by the triangle inequality $d(x, x_n) < d(x', x_n)$ and hence $F(x) \neq F(x')$. By Proposition 1.5.11 $F(X) \subset H$ is compact and by Proposition 1.5.13 F is a homeomorphism between X and $F(X)$. □

EXERCISE 2.7.2. Prove that the infinite-dimensioanl torus $\mathbb{T}^\infty$, the product of the countably many copies of the unit circle, has the same universality property as the Hilbert cube, that is, any compact separable metric space X is homeomorphic to a closed subset of $\mathbb{T}^\infty$.

2.7.5. Capacity and box dimension. For a compact metric space there is a notion of the "size" or capacity inspired by the notion of volume. Suppose X is a compact space with metric d. Then a set $E \subset X$ is said to be r-*dense* if $X \subset \bigcup_{x\in E} B_d(x,r)$, where $B_d(x,r)$ is the r-ball with respect to d around x (see **??**). Define the r-*capacity* of (X,d) to be the minimal cardinality $S_d(r)$ of an r-dense set.

For example, if $X=[0,1]$ with the usual metric, then $S_d(r)$ is approximately $1/2r$ because it takes over $1/2r$ balls (that is, intervals) to cover a unit length, and the $\lfloor 2+1/2r\rfloor$-balls centered at $ir(2-r)$, $0\le i\le \lfloor 1+1/2r\rfloor$ suffice. As another example, if $X=[0,1]^2$ is the unit square, then $S_d(r)$ is roughly r^{-2} because it takes at least $1/\pi r^2$ r-balls to cover a unit area, and, on the other hand, the $(1+1/r)^2$-balls centered at points (ir,jr) provide a cover. Likewise, for the unit cube $(1+1/r)^3$, r-balls suffice.

In the case of the ternary Cantor set with the usual metric we have $S_d(3^{-i})=2^i$ if we cheat a little and use closed balls for simplicity; otherwise, we could use $S_d((3-1/i)^{-i})=2^i$ with honest open balls.

One interesting aspect of capacity is the relation between its dependence on r [that is, with which power of r the capacity $S_d(r)$ increases] and dimension.

If $X=[0,1]$, then

$$\lim_{r\to 0} -\frac{\log S_d(r)}{\log r} \ge \lim_{r\to 0} -\frac{\log(1/2r)}{\log r} = \lim_{r\to 0}\frac{\log 2+\log r}{\log r}=1$$

and

$$\lim_{r\to 0} -\frac{\log S_d(r)}{\log r} \le \lim_{r\to 0} -\frac{\log\lfloor 2+1/2r\rfloor}{\log r} \le \lim -\frac{\log(1/r)}{\log r}=1,$$

so $\lim_{r\to 0} -\log S_d(r)/\log r = 1 = \dim X$. If $X=[0,1]^2$, then

$$\lim_{r\to 0} -\log S_d(r)/\log r = 2 = \dim X,$$

and if $X=[0,1]^3$, then

$$\lim_{r\to 0} -\log S_d(r)/\log r = 3 = \dim X.$$

This suggests that $\lim_{r\to 0} -\log S_d(r)/\log r$ defines a notion of dimension.

DEFINITION 2.7.11. If X is a totally bounded metric space (Definition 2.7.6), then

$$\operatorname{bdim}(X) := \lim_{r\to 0} -\frac{\log S_d(r)}{\log r}$$

is called the *box dimension* of X.

Let us test this notion on a less straightforward example. If C is the ternary Cantor set, then

$$\operatorname{bdim}(C) = \lim_{r\to 0} -\frac{\log S_d(r)}{\log r} = \lim_{n\to\infty} -\frac{\log 2^i}{\log 3^{-i}} = \frac{\log 2}{\log 3}.$$

If C_α is constructed by deleting a middle interval of relative length $1-(2/\alpha)$ at each stage, then $\operatorname{bdim}(C_\alpha) = \log 2/\log\alpha$. This increases to 1 as $\alpha \to 2$

(deleting ever smaller intervals), and it decreases to 0 as $\alpha \to \infty$ (deleting ever larger intervals). Thus we get a small box dimension if in the Cantor construction the size of the remaining intervals decreases rapidly with each iteration.

This illustrates, by the way, that the box dimension of a set may change under a homeomorphism, because these Cantor sets are pairwise homeomorphic. Box dimension and an associated but more subtle notion of *Hausdorff dimension* are the prime exhibits in the panoply of "fractal dimensions", the notion surrounded by a certain mystery (or mystique) at least for laymen. In the next section we will present simple calculations which shed light on this notion.

2.8. Metric spaces with symmetries and self-similarities

2.8.1. Euclidean space as an ideal geometric object and some of its close relatives. An outstanding, one may even say, the central, feature of Euclidean geometry, is an abundance of isometries in the Euclidean space. Not only there is isometry which maps any given point to any other point (e.g. the parallel translation by the vector connecting those points) but there are also isometries which interchange any given pair of points, e.g the central symmetry with respect to the midpoint of the interval connecting those points, or the reflection in the (hyper)plane perpendicular to that interval at the midpoint. The latter property distinguishes a very important class of Riemannian manifolds, called *symmetric spaces*. The next obvious examples of symmetric space after the Euclidean spaces are spheres $\mathbb{S}^n$ with the standard metric where the distance is measure along the shorter arcs of great circles. Notice that the metric induced from the embedding of $\mathbb{S}^n$ as the unit sphere into $\mathbb{R}^{n+1}$ also possesses all there isometries but the metric is not a Riemanninan metric, i.e. the distance cannot be calculated as the minimum of lengths of curves connecting two points, and thus this metric is much less interesting.

EXERCISE 2.8.1. How many isometries are there that interchange two points $x, y \in \mathbb{R}^n$ for different values of n?

EXERCISE 2.8.2. How many isometries are there that interchange two points $x, y \in \mathbb{S}^n$ for different values of n and for different configurations of points?

EXERCISE 2.8.3. Prove that the real projective space $\mathbb{R}P(n)$ with the metric inherited from the sphere (**??**) is a symmetric space.

EXERCISE 2.8.4. Prove that the torus $\mathbb{T}^n$ is with the metric inherited from $\mathbb{R}^n$ a symmetric space.

There is yet another remarkable property of Euclidean spaces which is not shared by other symmetric spaces: existence of *similarities*, i.e. transformations which preserve angles and changes all distances with the same coefficient of proportionality. It is interesting to point out that in the long quest to "prove" Euclid's fifth postulate, i.e. to deduce it from other axioms of Euclidean geometry, one among many equivalent formulations of the famous postulate is existence of a single pair of similar but not equal (not isometric) triangles. In the non-Euclidean hyperbolic geometry which results from adding the negation of the fifth postulates there no similar triangles and instead there is absolute unit of length! Incidentally the hyperbolic plane (as well as its higher-dimensional counterparts) is also a symmetric space. Existence of required symmetries can be deduced synthetically form the axioms common to Euclidean and non-Euclidean geometry, i.e. it belong s to so-called *absolute geometry*, the body of statement which can be proven in Euclidean geometry without the use of fifth postulate.

Metric spaces for which there exists a self-map which changes all distance with the same coefficient of proportionality different from one are called *self-similar*.

Obviously in a compact globally self-similar space which contain more one point the coefficient of proportionality for any similarity transformation must be less than one and such a transformation cannot be bijective; for non-compact spaces this is possible however.

2.8.2. Metrics on the Cantor set with symmetries and self-similarities. There is an interesting example of a similarity on the middle-third Cantor set, namely, $f_0\colon [0,1] \to [0,1]$, $f_0(x) = x/3$. Since f_0 is a contraction, it is also a contraction on every invariant subset, and in particular on the Cantor set. The unique fixed point is obviously 0. There is another contraction with the same contraction coefficient 1/3 preserving the Cantor set, namely $f_1(x) = \frac{x+2}{3}$ with fixed point 1. Images of these two contractions are disjoint and together they cover the whole Cantor set

EXERCISE 2.8.5. Prove that any similarity of the middle third Cantor set belongs to the semigroup generated by f_0 and f_1.

EXERCISE 2.8.6. Find infinitely many different self-similar Cantor sets on $[0,1]$ which contain both endpoints 0 and 1.

2.8.3. Other Self-Similar Sets. Let us describe some other interesting self-similar metric spaces that are of a different form. The *Sierpinski carpet* (see **??**) is obtained from the unit square by removing the "middle-ninth" square $(1/3,2/3)\times(1/3,2/3)$, then removing from each square $(i/3,i+1/3)\times(j/3,j+1/3)$ its "middle ninth," and so on. This construction can easily be described in terms of ternary expansion in a way that immediately suggests higher-dimensional analogs.

Another very symmetric construction begins with an equilateral triangle with the bottom side horizontal, say, and divide it into four congruent equilateral triangles of which the central one has a horizontal top side. Then one deletes this central triangle and continues this construction on the remaining three triangles. he resulting set is sometimes called *Sierpinski gasket*.

The *von Koch snowflake* is obtained from an equilateral triangle by erecting on each side an equilateral triangle whose base is the middle third of that side and continuing this process iteratively with the sides of the resulting polygon It is attributed to Helge von Koch (1904).

A three-dimensional variant of the Sierpinski carpet S is the Sierpinski sponge or Menger curve defined by $\{(x,y,z)\in[0,1]^3 \mid (x,y)\in S,\ (x,z)\in S\ (y,z)\in$

$S\}$. It is obtained from the solid unit cube by punching a 1/3-square hole through the center from each direction, then punching, in each coordinate direction, eight 1/9-square holes through in the right places, and so on. Both Sierpinski carper and Menger curve have important universality properties which we do not discuss in this book.

Let as calculate the box dimension of these new examples. For the square Sierpinski carpet we can cheat as in the capacity calculation for the ternary Cantor set and use closed balls (sharing their center with one of the small remaining cubes at a certain stage) for covers. Then $S_d(3^{-i}/\sqrt{2}) = 8^i$ and

$$\operatorname{bdim}(S) = \lim_{n\to\infty} -\frac{\log 8^i}{\log 3^{-i}/\sqrt{2}} = \frac{\log 8}{\log 3} = \frac{3\log 2}{\log 3},$$

which is three times that of the ternary Cantor set (but still less than 2, of course). For the triangular Sierpinski gasket we similarly get box dimension $\log 3/\log 2$.

The Koch snowflake K has $S_d(3^{-i}) = 4^i$ by covering it with (closed) balls centered at the edges of the ith polygon. Thus

$$\operatorname{bdim}(K) = \lim_{n\to\infty} -\frac{\log 4^i}{\log 3^{-i}} = \frac{\log 4}{\log 3} = \frac{2\log 2}{\log 3},$$

which is less than that of the Sierpinski carpet, corresponding to the fact that the iterates look much "thinner". Notice that this dimension exceeds 1, however, so it is larger than the dimension of a curve. All of these examples have (box) dimension that is not an integer, that is, fractional or "fractal". This has motivated calling such sets *fractals*.

Notice a transparent connection between the box dimension and coefficients of self-similarity on all self-similar examples.

2.9. Spaces of continuous maps

If X is a compact metrizable topological space (for example, a compact manifold), then the space $C(X,X)$ of continuous maps of X into itself possesses the C^0 or *uniform* topology. It arises by fixing a metric ρ in X and defining the distance d between $f, g \in C(X,X)$ by

$$d(f,g) := \max_{x\in X} \rho(f(x), g(x)).$$

The subset $\operatorname{Hom}(X)$ of $C(X,X)$ of homeomorphisms of X is neither open nor closed in the C^0 topology. It possesses, however, a natural topology as a complete metric space induced by the metric

$$d_H(f,g) := \max(d(f,g), d(f^{-1}, g^{-1})).$$

If X is σ-compact we introduce the compact–open topologies for maps and homeomorphisms, that is, the topologies of uniform convergence on compact sets.

We sometimes use the fact that equicontinuity gives some compactness of a family of continuous functions in the uniform topology.

THEOREM 2.9.1 (Arzelá–Ascoli Theorem). *Let X, Y be metric spaces, X separable, and $\mathcal{F}$ an equicontinuous family of maps. If $\{f_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ such that $\{f_i(x)\}_{i\in\mathbb{N}}$ has compact closure for every $x \in X$ then there is a subsequence converging uniformly on compact sets to a function f.*

Thus in particular a closed bounded equicontinuous family of maps on a compact space is compact in the uniform topology (induced by the maximum norm).

Let us sketch the proof. First use the fact that $\{f_i(x)\}_{i\in\mathbb{N}}$ has compact closure for every point x of a countable dense subset S of X. A diagonal argument shows that there is a subsequence f_{i_k} which converges at every point of S. Now equicontinuity can be used to show that for every point $x \in X$ the sequence $f_{i_k}(x)$ is Cauchy, hence convergent (since $\{f_i(x)\}_{i\in\mathbb{N}}$ has compact, hence complete, closure). Using equicontinuity again yields continuity of the pointwise limit. Finally a pointwise convergent equicontinuous sequence converges uniformly on compact sets.

EXERCISE 2.9.1. Prove that the set of Lipschitz real-valued functions on a compact metric space X with a fixed Lipschitz constant and bounded in absolute value by another constant is compact in $C(x, \mathbb{R})$.

EXERCISE 2.9.2. Is the closure in $C([0,1], \mathbb{R})$ (which is usually denoted simply by $C([0,1])$) of the set of all differentiable functions which derivative bounded by 1 in absolute value and taking value 0 at 1/2 compact?

2.10. Spaces of closed subsets of a compact metric space

2.10.1. Hausdorff distance: definition and compactness. An interesting construction in the theory of compact metric spaces is that of the Hausdorff metric:

DEFINITION 2.10.1. If (X, d) is a compact metric space and $K(X)$ denotes the collection of closed subsets of X, then the *Hausdorff metric* d_H on $K(X)$ is defined by

$$d_H(A, B) := \sup_{a\in A} d(a, B) + \sup_{b\in B} d(b, A),$$

where $d(x, Y) := \inf_{y\in Y} d(x, y)$ for $Y \subset X$.

Notice that d_H is symmetric by construction and is zero if and only if the two sets coincide (here we use that these sets are closed, and hence compact, so the "sup" are actually "max"). Checking the triangle inequality requires a little extra work. To show that $d_H(A, B) \le d_H(A, C) + d_H(C, B)$, note that $d(a, b) \le d(a, c) + d(c, b)$ for $a \in A$, $b \in B$, $c \in C$, so taking the infimum over b we get $d(a, B) \le d(a, c) + d(c, B)$ for $a \in A$, $c \in C$. Therefore, $d(a, B) \le d(a, C) + \sup_{c\in C} d(c, B)$ and $\sup_{a\in A} d(a, B) \le \sup_{a\in A} d(a, C) + \sup_{c\in C} d(c, B)$. Likewise, one gets $\sup_{b\in B} d(b, A) \le \sup_{b\in B} d(b, C) + \sup_{c\in C} d(c, A)$. Adding the last two inequalities gives the triangle inequality.

PROPOSITION 2.10.2. *The Hausdorff metric on the closed subsets of a compact metric space defines a compact topology.*

PROOF. We need to verify total boundedness and completeness. Pick a finite $\epsilon/2$-net N. Any closed set $A \subset X$ is covered by a union of ϵ-balls centered at points of N, and the closure of the union of these has Hausdorff distance at most ϵ from A. Since there are only finitely many such sets, we have shown that this metric is totally bounded. To show that it is complete, consider a Cauchy sequence (with respect to the Hausdorff metric) of closed sets $A_n \subset X$. If we let $A := \bigcap_{k\in\mathbb{N}} \overline{\bigcup_{n\geq k} A_n}$, then one can easily check that $d(A_n, A) \to 0$. □

EXERCISE 2.10.1. Prove that for the Cantor set C the space $K(C)$ is homeomorphic to C.

EXERCISE 2.10.2. Prove that $K([0, 1])$ contains a subset homeomorphic to the Hilbert cube.

2.10.2. Existence of a minimal set for a continuous map. Any homeomorphism of a compact metric space X induces a natural homeomorphism of the collection of closed subsets of X with the Hausdorff metric, so we have the following:

PROPOSITION 2.10.3. *The set of closed invariant sets of a homeomorphism f of a compact metric space is a closed set with respect to the Hausdorff metric.*

PROOF. This is just the set of fixed points of the induced homeomorphism; hence it is closed. □

We will now give a nice application of the Hausdorff metric. Brouwer fixed point Theorem does not extend from the disc to continuous maps of other spaces even as simple and and nice as the circle. The simplest example of a continuous map (in fact a self–homeomorphism) which does not have have fixed points is a rotation of the circle; if the angle of rotation is a rational multiple of π all points are periodic with the same period; otherwise there are no periodic points.

However, there is a nice generalization which works for any compact Hausdorff spaces. An obvious property of a fixed or a periodic point for a continuous map is its minimality: it is an invariant closed set which has no invariant subsets.

DEFINITION 2.10.4. An invariant closed subset A of a continuous map $f\colon X \to X$ is *minimal* if there are no nonempty closed f-invariant subsets of A.

THEOREM 2.10.5. *Any continuous map f of a compact Hausdorff space X with a countable base into itself has an invariant minimal set.*

PROOF. By Corollary 2.6.2 the space X is metrizable. Fix a metric d on X and consider the Hausdorff metric on the space $K(X)$ of all closed subsets of X. Since any closed subset A of X is compact (Proposition 1.5.2) $f(A)$ is also compact (Proposition 1.5.11) and hence closed (Corollary 2.6.2). Thus f naturally induces a map $f_*\colon K(X) \to K(X)$ by setting $f_*(A) = f(A)$. A direct calculation shows that the map f_* is continuous in the topology induced by the Hausdorff metric. Closed f-invariant subsets of X are fixed points of f_*. The set of all such sets is closed, hence compact subset $I(f)$ of $K(X)$. Consider for each $B \in I(f)$ all $A \in I(f)$ such that $A \subset B$. Such A form a closed, hence compact, subset $I_B(f)$. Hence the function on $I_B(f)$ defined by $d_H(A, B)$ reaches its maximum, which we denote by $m(B)$, on a certain f-invariant set $M \subset B$.

Notice that the function $m(B)$ is also continuous in the topology of Hausdorff metric. Hence it reaches its minimum m_0 on a certain set N. If $m_0 = 0$, the set N is a minimal set. Now assume that $m_0 > 0$.

Take the set $M \subset B$ such that $d_H(M, B) = m(B) \geq m_0$. Inside M one can find an invariant subset M_1 such that $d_H(M_1, M) \geq m_0$. Notice that since $M_1 \subset M,\ d_H(M_1, B) \geq d_H(M, B) = m(B) \geq m_0$.

Continuing by induction we obtain an infinite sequence of nested closed invariant sets $B \supset M \supset M_1 \supset M_2 \supset \cdots \supset M_n \supset \ldots$ such that the Hausdorff distance between any two of those sets is at least m_0. This contradicts compactness of $K(X)$ in the topology generated by the Hausdorff metric. □

EXERCISE 2.10.3. Give detailed proofs of the claims used in the proof of Theorem 2.10.5:

- the map $f_* \colon K(X) \to K(X)$ is continuous;
- the function $m(\cdot)$ is continuous;
- $d_H(M_i, M_j) \geq m_0$ for $i, j = 1, 2, \ldots; i \neq j$.

EXERCISE 2.10.4. For every natural number n give an example of a homeomorphism of a compact path connected topological space which has no fixed points and has exactly n minimal sets.

2.11. Topological groups

In this section we introduce groups which carry a topology invariant under the group operations. A *topological group* is a group endowed with a topology with respect to which all *left translations* $L_{g_0} \colon g \mapsto g_0 g$ and *right translations* $R_{g_0} \colon g \mapsto g g_0$ as well as $g \mapsto g^{-1}$ are homeomorphisms. Familiar examples are $\mathbb{R}^n$ with the additive structure as well as the circle or, more generally, the n-torus, where translations are clearly diffeomorphisms, as is $x \mapsto -x$.

2.12. Problems

EXERCISE 2.12.1. Prove that every metric space is homeomorphic to a bounded space.

EXERCISE 2.12.2. Prove that in a compact set A in metric space X there exists a pair or points $x, y \in A$ such that $d(x, y) = \operatorname{diam} A$.

EXERCISE 2.12.3. Suppose a function $d \colon X \times X \to \mathbb{R}$ satisfies conditions (2) and (3) of Definition 2.1.1 but not (1). Find a natural way to modify this function so that the modified function becomes a metric.

EXERCISE 2.12.4. Let S be a smooth surface in $\mathbb{R}^3$, i.e. it may be a non-critical level of a smooth real-valued function, or a closed subset locally given as a graph when one coordinate is a smooth function of two others. S carries two metrics: (i) induced from $\mathbb{R}^3$ as a subset of a metric space, and (ii) the natural internal distance given by the minimal length of curves in S connecting two points.

Prove that if these two metrics coincide then S is a plane.

EXERCISE 2.12.5. Introduce a metric d on the Cantor set C (generating the Cantor set topology) such that (C, d) cannot be isometrically embedded to $\mathbb{R}^n$ for any n.

EXERCISE 2.12.6. Introduce a metric d on the Cantor set C such that (C, d) is not Lipschitz equivalent to a subset of $\mathbb{R}^n$ for any n.

EXERCISE 2.12.7. Prove that the set of functions which are not Hölder continuous at any point is a residual subset of $C([0, 1])$.

EXERCISE 2.12.8. Let $f\colon [0, 1] \to \mathbb{R}^2$ be α-Höder with $\alpha > 1/2$. Prove that $f([0, 1)]$ is nowhere dense.

EXERCISE 2.12.9. Find a generalization of the previous statement for the maps of the m-dimensional cube I^m to $\mathbb{R}^n$ with $m < n$.

EXERCISE 2.12.10. Prove existence of 1/2-Hölder surjective map $f\colon [0, 1] \to I^2$. (Such a map is usually called a *Peano curve*).

EXERCISE 2.12.11. Find a Riemannian metric on the complex projective space $\mathbb{C}P(n)$ which makes it a symmetric space.

EXERCISE 2.12.12. Prove that $\mathbb{S}^n$ is not self-similar.