An introduction to analysis

Swapneel Mahajan

Department of Mathematics
Indian Institute of Technology Mumbai
Powai, Mumbai 400 076
India
E-mail address: swapneel@math.iitb.ac.in
URL: http://www.math.iitb.ac.in/~swapneel

Contents

Prerequisites.

- You should be familiar with the basic language of set theory.
- You should be very comfortable with one-variable calculus, including the ϵ-δ definition for the convergence of real sequences, and similarly of continuous maps.
- You should be willing to work hard to fill in gaps in your understanding.

References. Here is a list of useful references, which is by no means exhaustive.

- For analysis, rudin [21], pugh [19], browder [8], munkres [18]. Also useful are apostol [1], simmons [22], tao [24, 25].
- For algebra, artin [3], dummit-foote [10].
- For set theory, halmos [11], munkres [18, Chapter 1].
- For posets, davey-priestley [9].
- For category theory, maclane [16].
- Wikipedia is a good online source for getting a birds-eye-view of many of the concepts discussed in these notes. Blogs are also useful.

Pick a book that suits you. To understand the subject matter, it is not necessary to understand each and every sentence written in a particular book.

A number of exercises are included in the notes (many with complete solutions).

Lecture-wise content.

(1) Abstract categories. The category of sets. Tabulate some important categories.
(2) Relations on a set. Highlight equivalence relations.
(3) Introduce the category of posets.
(4) Define a functor, and illustrate with subcategories. Explain the forgetful functor from posets to sets.
(5) Monoids and groups.
(6) Rings, fields, vector spaces.
(7) Cauchy sequences, and the construction of $\mathbb{R}$. Show that $\mathbb{R}$ is a field.
(8) Define a linear order on $\mathbb{R}$ and show that it is an ordered field.
(9) Binary expansions.
(10) The least upper bound property and existence of nth roots.
(11) Sequences of real numbers and the completeness property.
(12) Continuous functions on the real line and the intermediate value theorem.
(13) Metric spaces.
(14) Convergence and Cauchyness of sequences in a metric space.
(15) The completion functor.
(16) Topological spaces. The functor from metric spaces to topological spaces.
(17) Full and faithful functors. Closure operators and closed sets.
(18) Hausdorff spaces, subspace topology, compact topological spaces, closed interval is compact.
(19) Bounded metric spaces, compact subspaces of a metric space are closed and bounded.
(20) Product topology, product of compact sets.
(21) Finite products in a category, Heine-Borel Theorem.
(22) Compactness and continuity, compactness in terms of closed sets.
(23) Limit point compactness and sequential compactness.
(24) Characterizations of compactness for metric spaces.

(25) Connected topological spaces.
(26) Normed linear spaces.
(27) Review of categories, in particular, of the notion of isomorphism. Categories with one object and monoids. Isomorphism-closed subcategories.
(28) Adjoint functors. Construction of left adjoints of the forgetful functors to sets from posets, topological spaces, monoids, vector spaces.
(29) The road ahead.

Future reading.

- For topology, munkres [18], armstrong [2], janich [12, 7], warner [26],
- For differentiable manifolds and differential geometry, lee [14], boothby [6], morita [17], spivak [23],
- For measure theory, rudin [21], royden [20]
- For probability theory, billingsley [5],
- For functional analysis, limaye [15]

CHAPTER 1

Getting oriented

We discuss sets, relations on sets with particular emphasis on equivalence relations and partial orders, algebraic systems such as monoids, groups, rings, fields, vector spaces. The categorical viewpoint is emphasized, and the relationships among these notions are explained via functors.

1.1. Categories and functors

This section is to be read in parallel with the subsequent sections. We freely refer to various examples which will only be defined later in the text.

Mathematics speaks the language of category theory. Some basic categories which routinely appear in mathematics are listed in Table 1.1.

TABLE 1.1. Some basic categories of mathematics.

Category	Objects	Morphisms
Set	sets	functions
Poset	partially ordered sets	order-preserving maps
Lin	linearly ordered sets	order-preserving maps
Metric	metric spaces	continuous maps
Top	topological spaces	continuous maps
Monoid	monoids	monoid homomorphisms
Group	groups	group homomorphisms
Ring	rings	ring homomorphisms
Field	fields	field extensions
$\mathsf{Vec}_{\Bbbk}$	vector spaces over a field $\Bbbk$	$\Bbbk$-linear maps
NLS	normed linear spaces	bounded linear maps

The category of sets consists of

- all sets $X, Y, \dots$ (possibly infinite),
- all functions $X \to Y$ for any sets X and Y.

We denote this category by Set. Set theory is the study of this category.

The other entries in the table are to be interpreted in a similar manner. For example, $\mathsf{Vec}_{\Bbbk}$ is the category whose objects are vector spaces over $\Bbbk$ and whose morphisms are $\Bbbk$-linear maps. Linear algebra is the study of this category.

1.1.1. Categories. A category C consists of the following data.

- objects $a, b, \ldots,$
- for any two objects a and b, a set of morphisms $\mathsf{C}(a, b)$. We denote an element of $\mathsf{C}(a, b)$ by $f : a \to b$, and say that f is a morphism from a to b.
- for each object a, there is a distinguished element of $\mathsf{C}(a, a)$ denoted id_a.
- for any three objects a, b, c, a binary operation (function)

$$\mathsf{C}(a, b) \times \mathsf{C}(b, c) \to \mathsf{C}(a, c), \qquad (f, g) \mapsto g \circ f,$$

  which is associative and unital:
  If $f : a \to b$, $g : b \to c$, $h : c \to d$, then $h \circ (g \circ f) = (h \circ g) \circ f$.
  If $f : a \to b$, then $\mathrm{id}_b \circ f = f = f \circ \mathrm{id}_a$.

We refer to the binary operation above as the composition rule. We say that $g \circ f$ is the composite of f and g, or is obtained by composing f with g. The set $\mathsf{C}(a, b)$ could be empty, that is, there are no morphisms from a to b.

We say a morphism $f : a \to b$ is an *isomorphism* in C if there exists a morphism $g : b \to a$ such that $f \circ g = \mathrm{id}_b$ and $g \circ f = \mathrm{id}_a$. The morphism g is called the inverse of f. We say that the objects a and b are *isomorphic* if there exists an isomorphism $f : a \to b$. Two objects can be isomorphic via many different isomorphisms.

An isomorphism from an object to itself is usually called an *automorphism.*

A category can be pictured as a graph with objects as vertices, and morphism as arrows (directed edges). The identity morphisms are loops. The composition rule is illustrated by the diagram

$$\begin{tikzcd} & b \arrow[dr, "g"] & \\ a \arrow[ur, "f"] \arrow[rr, "g \circ f"'] & & c. \end{tikzcd}$$

This is an instance of a 'commutative' diagram.

1.1.2. Finite categories. Here are some concrete 'tiny' categories.

- the *empty category* with no objects and no morphisms. In a way, there is nothing to say.
- the category with one object a and one morphism id_a. There is only one way to define the composition rule. This is also called the *one-arrow category.*

$$\begin{tikzcd} a \arrow[loop above, "\mathrm{id}_a"] \end{tikzcd}$$

- the category with two objects a and b, exactly one morphism from a to b, and the identity morphisms id_a and id_b. This is also called the *interval category.*

$$\begin{tikzcd} a \arrow[loop left, "\mathrm{id}_a"] \arrow[r] & b \arrow[loop right, "\mathrm{id}_b"] \end{tikzcd}$$

- There is an even smaller category on two objects a and b consisting of only the identity morphisms id_a and id_b.

$$\begin{tikzcd} a \arrow[loop left, "\mathrm{id}_a"] & b \arrow[loop right, "\mathrm{id}_b"] \end{tikzcd}$$

  This is the *discrete category* on two objects. One can define the discrete category on any set of objects: the only morphisms are the identity morphisms. Note: The one-arrow category is the discrete category on one object.

1.1.3. Products and coproducts. Fix a category C. An object a in C is an *initial object* if for any object b, there is a unique morphism $a \to b$. A category may not have any initial objects or it may have more than one initial object. All initial objects are isomorphic. (See exercise.)

Dually, an object a in C is a *terminal object* if for any object b, there is a unique morphism $b \to a$. Similar remarks apply.

Given objects a and b in C, their *product* is an object $a \times b$ in C equipped with morphisms $a \times b \to a$ and $a \times b \to b$ such that the following universal property holds. For any object c and morphisms $f : c \to a$ and $g : c \to b$, there is a unique morphism $c \to a \times b$ such that the diagram

$$\begin{tikzcd}
& c \arrow[dl, "f"'] \arrow[d, dashed] \arrow[dr, "g"] & \\
a & a \times b \arrow[l] \arrow[r] & b
\end{tikzcd}$$

commutes.

Dually, given objects a and b in C, their *coproduct* is an object $a \sqcup b$ in C equipped with morphisms $a \to a \sqcup b$ and $b \to a \sqcup b$ such that the following universal property holds. For any object c and morphisms $f : a \to c$ and $g : b \to c$, there is a unique morphism $a \sqcup b \to c$ such that the diagram

$$\begin{tikzcd}
& c & \\
a \arrow[ur, "f"] \arrow[r] & a \sqcup b \arrow[u, dashed] & b \arrow[l] \arrow[ul, "g"']
\end{tikzcd}$$

commutes.

The (co)product of a and b may not exist, but if it exists, then it is unique up to isomorphism. (See exercise.)

More generally, given objects $a_1, \dots, a_n$, their product is an object $\prod_{i=1}^n a_i$ equipped with morphisms $\prod_{i=1}^n a_i \to a_i$, one for each i such that the following universal property holds. For any object c and morphisms $f_i : c \to a_i$ one for each i, there is a unique morphism $c \to \prod_{i=1}^n a_i$ such that the diagram

$$\begin{tikzcd}
c \arrow[d, dashed] \arrow[dr, "f_i"] & \\
\prod_{i=1}^n a_i \arrow[r] & a_i
\end{tikzcd}$$

commutes for each i.

A finite coproduct is defined dually by reversing arrows.

1.1.4. Subcategories. A subcategory of a category C is a category whose objects and morphisms are picked from the objects and morphisms of C, with the composition rule and identities inherited from C. For instance, consider the category whose objects are finite sets, and morphisms are functions. This is a subcategory of the category of sets. Here we have restricted the class of objects and kept all morphisms between the allowed objects. This is called a *full subcategory*. Other such examples are finite-dimensional vector spaces inside vector spaces, finite groups inside groups, abelian groups inside groups, linearly ordered sets inside posets, compact topological spaces inside topological spaces, and so on. In contrast, we could take the category whose objects are sets, but whose morphisms are injective maps. This is also a subcategory of the category of sets. Here we have kept all the objects but restricted the class of morphisms. In general, in a subcategory, we restrict both the objects and the morphisms.

A subcategory D of C is called *isomorphism-closed* if the following condition holds. If a is an object in D and $f : a \to b$ is an isomorphism in C, then b is an object in D and f is a morphism in D. All examples of subcategories mentioned above are isomorphism-closed.

1.1.5. Functors. A functor is a way to compare categories. Let C and D be two categories. A functor $\mathcal{F} : \mathsf{C} \to \mathsf{D}$ assigns to every object a in C, an object $\mathcal{F}(a)$ in D, and to every morphism $f : a \to b$ in C a morphism $\mathcal{F}(f) : \mathcal{F}(a) \to \mathcal{F}(b)$ in D such that

$$\mathcal{F}(g \circ f) = \mathcal{F}(g) \circ \mathcal{F}(f) \qquad \text{and} \qquad \mathcal{F}(\mathrm{id}) = \mathrm{id}.$$

Observe that the objects in the categories listed in Table 1.1 are sets at the very least. They are sets equipped with some extra structure. Similarly, the morphisms are functions at the very least. This says that there is a functor from any of these categories to Set. Such a functor is called a forgetful functor. For instance: A ring has more structure than a monoid, in fact, there are two monoids present in a ring, so there are two forgetful functors $\mathsf{Ring} \to \mathsf{Monoid}$.

One must understand the difference between imposing structure and imposing conditions. The first gives rise to forgetful functors and the second to inclusion functors, that is, subcategories. For instance, being a group is a property of a monoid. So $\mathsf{Group} \to \mathsf{Monoid}$ is an inclusion functor. Similarly, being a field is a property of a ring. So $\mathsf{Field} \to \mathsf{Ring}$ is also an inclusion functor.

There is a related notion of a contravariant functor. It assigns to every object a in C, an object $\mathcal{F}(a)$ in D, and to every morphism $f : a \to b$ in C, a morphism $\mathcal{F}(f) : \mathcal{F}(b) \to \mathcal{F}(a)$ in D which respects composition. Note the reversal of the arrow. By making use of what is called an opposite category, a contravariant functor can be interpreted as a usual functor.

Suppose $\mathcal{F} : \mathsf{C} \to \mathsf{D}$ is a functor. Then for any objects a and b of C, there is a function

$$\mathsf{C}(a, b) \to \mathsf{D}(\mathcal{F}(a), \mathcal{F}(b)), \qquad f \mapsto \mathcal{F}(f).$$

We say that $\mathcal{F}$ is *faithful* if this function is injective, *full* if it is surjective, and *full and faithful* if it is bijective, for all a and b.

Observe that any inclusion functor (arising from a subcategory) is faithful. Further if the subcategory is full, then the inclusion functor is full and faithful.

1.1.6. Natural transformation. Suppose C and D are categories, and $\mathcal{F}$ and $\mathcal{G}$ are functors both from C to D. A natural transformation $\eta : \mathcal{F} \Rightarrow \mathcal{G}$ assigns to each object a of C a morphism

$$\eta_a : \mathcal{F}(a) \to \mathcal{G}(a)$$

such that for any morphism $f : a \to b$ in C, the diagram

$$\begin{tikzcd}
\mathcal{F}(a) \arrow[r, "\mathcal{F}(f)"] \arrow[d, "\eta_a"'] & \mathcal{F}(b) \arrow[d, "\eta_b"] \\
\mathcal{G}(a) \arrow[r, "\mathcal{G}(f)"'] & \mathcal{G}(b)
\end{tikzcd}$$

commutes.

We say $\mathcal{F}$ and $\mathcal{G}$ are *naturally isomorphic* if there exists a natural transformation η between them for which all η_a are isomorphisms in D.

1.1.7. Equivalence of categories. Two categories C and D are called *equivalent* if there exist functors

$$\mathcal{F}\colon \mathsf{C} \to \mathsf{D} \quad \text{and} \quad \mathcal{G}\colon \mathsf{D} \to \mathsf{C}$$

such that $\mathcal{G}\mathcal{F}$ is naturally isomorphic to the identity functor on C, and $\mathcal{F}\mathcal{G}$ is naturally isomorphic to the identity functor on D. In this situation, $\mathcal{F}$ (or $\mathcal{G}$) is said to be an *equivalence* between C and D.

A full and faithful functor $\mathcal{F} : \mathsf{C} \to \mathsf{D}$ induces an equivalence of categories between C and the full subcategory of D given by the image of $\mathcal{F}$.

1.1.8. Adjoint functors. Let $\mathcal{F}\colon \mathsf{C} \to \mathsf{D}$ and $\mathcal{G}\colon \mathsf{D} \to \mathsf{C}$ be a pair of functors. We say that $\mathcal{F}$ is a *left adjoint* to $\mathcal{G}$ or that $\mathcal{G}$ is a *right adjoint* to $\mathcal{F}$ if for each object a in C and x in D, there exists a bijection

$$\mathsf{D}(\mathcal{F}(a), x) \xrightarrow{\cong} \mathsf{C}(a, \mathcal{G}(x)) \tag{1.1}$$

which is natural in a and x. The term "natural" means that for any morphism $a \to a'$ in C and $x \to x'$ in D, the diagrams

$$\begin{tikzcd} \mathsf{D}(\mathcal{F}(a), x) \arrow[r] & \mathsf{C}(a, \mathcal{G}(x)) \\ \mathsf{D}(\mathcal{F}(a'), x) \arrow[r] \arrow[u] & \mathsf{C}(a', \mathcal{G}(x)) \arrow[u] \end{tikzcd} \qquad \begin{tikzcd} \mathsf{D}(\mathcal{F}(a), x) \arrow[r] \arrow[d] & \mathsf{C}(a, \mathcal{G}(x)) \arrow[d] \\ \mathsf{D}(\mathcal{F}(a), x') \arrow[r] & \mathsf{C}(a, \mathcal{G}(x')) \end{tikzcd}$$

commute. The horizontal arrows are instances of (1.1). The vertical arrows are induced from the composition law and the functors.

A left (right) adjoint of a given functor may or may not exist, but if it does exist, then it is unique up to a natural isomorphism. That is, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are both left adjoints of a functor $\mathcal{G}$, then $\mathcal{F}_1$ and $\mathcal{F}_2$ are naturally isomorphic.

Problems.

(1) Suppose C is a category with identity morphisms denoted id_a. Suppose for each object a, we are given a morphism id'_a which satisfies the unitality axiom. Then show that $\mathrm{id}_a = \mathrm{id}'_a$.
(2) Show that any isomorphism $f : a \to b$ in a category C has a unique inverse.
(3) Show that any two initial (terminal) objects in a category are isomorphic.
(4) Let C be a category and a and b be any two objects. Show that the (co)product of a and b is unique up to isomorphism (assuming that it exists).
(5) Suppose $f : a \to b$ is an isomorphism. Then for any object c, show that there is a bijection
$$\mathsf{C}(b, c) \to \mathsf{C}(a, c), \qquad h \mapsto h \circ f.$$
(6) Show that: A functor preserves isomorphisms. That is, if $\mathcal{F} : \mathsf{C} \to \mathsf{D}$ is a functor and $f : a \to b$ is an isomorphism in C, then $\mathcal{F}(f) : \mathcal{F}(a) \to \mathcal{F}(b)$ is an isomorphism in D.
(7) Let C be a category. What are all functors from the one-arrow category to C? What are all functors from the interval category to C?

(8) Suppose C is a category. Pick any object a in C. Consider the gadget with one object a, and set of morphisms $\mathsf{C}(a,a)$, with binary operation

$$\mathsf{C}(a,a) \times \mathsf{C}(a,a) \to \mathsf{C}(a,a)$$

induced from the composition rule of C. Does this define a subcategory of C? What are categories with exactly one object? What are functors between categories with exactly one object?

What happens if instead of picking $\mathsf{C}(a,a)$ you only pick the subset of all automorphisms of a? Do you get a subcategory? What are categories with exactly one object in which every morphism is an isomorphism? What about functors between such categories?

(9) A *groupoid* is a category in which every morphism is an isomorphism. What is a groupoid with one object?

(10) A category is called *indiscrete* if there is a unique morphism from one object to any other object. Show that any indiscrete category is a groupoid.

(11) Show that the composite of two functors is a functor. For any category C, one always has the identity functor which takes every object to itself, and every morphism to itself. Does it make sense to define a category whose objects are categories, and morphisms are functors?

1.2. Category of sets

I assume that you have a working knowledge of sets and functions. In particular, that you are familiar with

- injective (into), surjective (onto), and bijective (one-one) functions.
- standard operations on sets such as unions, intersections, complements, ... Do you understand the meaning of $\cup_{i\in I} A_i$?

1.2.1. Category of sets. Let Set denote the category whose objects are sets, and morphisms are functions. For every set X, we have the identity function

$$\mathrm{id}_X : X \to X, \qquad \mathrm{id}_X(x) = x.$$

These are the identity morphisms.

Given a function $f : X \to Y$ and a function $g : Y \to Z$, there is a function $g \circ f : X \to Z$ defined by

$$(g \circ f)(x) = g(f(x)).$$

This is the composition rule for morphisms. It is associative and unital:

$$(h\circ(g\circ f))(x) = h(g(f(x))) = ((h\circ g)\circ f)(x) \quad \text{and} \quad (\mathrm{id}_b\circ f)(x) = f(x) = (f\circ \mathrm{id}_a)(x).$$

Thus we obtain a category.

Note that $f : A \to B$ is a bijection iff there exists a function $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. Thus a bijection is the same as an isomorphism in the category of sets. Two sets are isomorphic iff there is a bijection between them. Note that there could be many bijections between two given sets.

1.2.2. Number systems. The set of natural numbers is

$$\mathbb{N} := \{0, 1, 2, 3 \dots\}.$$

Note that 0 is included in this list. Is it not as "natural" as any of the remaining numbers? We will write $\mathbb{N}_+$ for the subset of natural numbers which excludes 0.

The set of integers is

$$\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3 \dots\}.$$

Do you understand what -1 really is? The set of rationals is

$$\mathbb{Q} := \{\frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\}.$$

The notation $\frac{m}{n}$ requires some explanation. You can think of it as a new symbol, and two symbols $\frac{m}{n}$ and $\frac{m'}{n'}$ are the same iff $mn' = m'n$. A formal way to understand this is through equivalence relations as we will soon see.

1.2.3. Initial object, terminal object, product, coproduct. The category of sets has a unique initial object which is the empty set $\emptyset$. Any singleton set is a terminal object. (Note that all singleton sets are isomorphic.)

Given sets X and Y, their *cartesian product* is defined by

$$X \times Y := \{(x, y) \mid x \in X, y \in Y\}.$$

Note that there are canonical projection maps

$$X \times Y \to X, \quad (x, y) \mapsto x, \quad \text{and} \quad X \times Y \to Y, \quad (x, y) \mapsto y.$$

Cartesian product satisfies the following universal property. Given any set Z and functions $f : Z \to X$ and $g : Z \to Y$, there is a unique function $Z \to X \times Y$ denoted (f, g) such that the diagram

$$\begin{tikzcd} & Z \arrow[dl, "f"'] \arrow[d, dashed] \arrow[dr, "g"] & \\ X & X \times Y \arrow[l] \arrow[r] & Y \end{tikzcd}$$

commutes. (Can you see that $(f, g)(z) := (f(z), g(z))$ makes the diagram commute, and that there is no other choice?) Conclusion: Cartesian product is the product in the category of sets.

Given sets X and Y, their *disjoint union* is defined by

$$X \sqcup Y := \{(x, 0) \mid x \in X\} \cup \{(y, 1) \mid y \in Y\}.$$

In the rhs, the first set is a copy of X, while the second is a copy of Y, but these copies have no common elements (though X and Y may have common elements). Do you understand the difference between $X \sqcup Y$ and $X \cup Y$? Note that there are canonical inclusion maps

$$X \to X \sqcup Y, \quad x \mapsto (x, 0), \quad \text{and} \quad Y \to X \sqcup Y, \quad y \mapsto (y, 1).$$

The disjoint union satisfies the following universal property. Given any set Z and functions $f : X \to Z$ and $g : Y \to Z$, there is a unique function $X \sqcup Y \to Z$ denoted $f \sqcup g$ such that the diagram

$$\begin{tikzcd} & Z & \\ X \arrow[ur, "f"] \arrow[r] & X \sqcup Y \arrow[u, dashed] & Y \arrow[ul, "g"'] \arrow[l] \end{tikzcd}$$

commutes. Conclusion: Disjoint union is the coproduct in the category of sets. Why does this universal property fail if we use $X \cup Y$ instead of $X \sqcup Y$.

1.2.4. Relations on a set. Let X be any set. A *relation* on X is a subset of $X \times X$.

Suppose R is a relation and $(x, y) \in R$. Alternative notations for denoting this are $x \sim_R y$ and xRy. We read this as "x is related to y under R".

One can visualize a relation as a graph whose vertices are the elements of X, and there is an arrow (edge) from x to y whenever $(x, y) \in R$. Does this remind you of a category?

A relation R is

- reflexive if $(x, x) \in R$ for all $x \in X$.
- irreflexive if $(x, x) \notin R$ for all $x \in X$.
- symmetric if $(x, y) \in R$ implies $(y, x) \in R$ for all $x, y \in X$.
- antisymmetric if $(x, y) \in R$ and $(y, x) \in R$ implies $x = y$ for all $x, y \in X$.
- transitive if $(x, y) \in R$ and $(y, z) \in R$ implies $(x, z) \in R$ for all $x, y, z \in X$.

Each of these is a property that a relation either has or does not have.

A *preposet* is a set equipped with a relation which is transitive and reflexive, a *poset* is a set equipped with a relation which is transitive, reflexive, antisymmetric, an *equivalence relation* is a relation which is transitive, symmetric, reflexive, and a *graph without loops or multiple edges* is a relation which is symmetric and irreflexive. Ponder over the last definition. These are summarized in Table 1.2.

TABLE 1.2. Relations.

Name	Type of relation
relations	arbitrary
preposets	transitive and reflexive
posets	transitive, antisymmetric, reflexive
equivalence relations	transitive, symmetric, reflexive
graphs without loops or multiple edges	symmetric and irreflexive

A *named set* is a triple (X, A, Y), where X and Y are any sets and A is a subset of $X \times Y$. This concept very similar to that of a relation has emerged recently in many areas.

1.2.5. Equivalence relations. Recall that an equivalence relation on a set X is a relation which is transitive, symmetric, reflexive. In this context, apart from $x \sim_R y$ and xRy, the notation $x \equiv_R y$ is also used. We read this as "x is equivalent to y under R". Sometimes R is dropped from the notation, or we even think of $\sim$ as the relation and then write $x \sim y$. In this notation, the definition of an equivalence relation takes the following form. For all $x, y, z \in X$,

- $x \sim x$,
- $x \sim y$ implies $y \sim x$,
- $x \sim y$ and $y \sim z$ implies $x \sim z$.

The *equivalence class* of an element x under $\sim$, denoted $[x]$, is defined as

$$[x] := \{y \in X \mid x \sim y\}.$$

For any $x, y \in X$, if $x \sim y$, then $[x] = [y]$, else $[x]$ and $[y]$ are disjoint (that is, their intersection is the empty set). Why is this true? Let

$$X_\sim := \{[x] \mid x \in X\}.$$

This is the set whose elements are the equivalence classes of $\sim$. Note that there is a canonical quotient map

$$X \twoheadrightarrow X_\sim, \qquad x \mapsto [x].$$

Observation 1.1. A function $f : X \to Y$ factors through this quotient map to yield a commutative diagram

$$\begin{tikzcd} X \arrow[d, two heads] \arrow[dr, "f"] & \\ X_\sim \arrow[r, dotted] & Y \end{tikzcd}$$

iff $f(x) = f(y)$ whenever $x \sim y$.

(The usage of factors through should remind you of factorization. Here we are 'factorizing' f by writing it as a composite, with one of the factors being the quotient map.)

Example 1.2. Consider the following equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$:

$$(m, n) \sim (m', n') \iff mn' = nm'.$$

Check that this is an equivalence relation. Transitivity requires some thought. An equivalence class is a rational number, and the set of all equivalence classes $(\mathbb{Z} \times \mathbb{Z})_\sim$ is the set of rational numbers $\mathbb{Q}$. We write $\frac{m}{n}$ instead of (m, n) but this is by tradition. The function

$$\mathbb{Z} \hookrightarrow \mathbb{Q}, \qquad m \mapsto \frac{m}{1}$$

is injective, and in this manner, one can think of $\mathbb{Z}$ as a subset of $\mathbb{Q}$. Such a construction can be carried out in the general context of commutative rings and is termed localization [4, Chapter 3].

Problems.

(1) Let $f : X \to Y$ be a function between sets X and Y. Define for $A \subseteq X$ and for $B \subseteq Y$,

$$f(A) := \{y \in Y \mid f(x) = y \text{ for some } x \in A\}$$
$$f^{-1}(B) := \{x \in X \mid f(x) = y \text{ for some } y \in B\}.$$

(We emphasize that $f^{-1}(B)$ is a notation, and we are not assuming here that f is a bijection. This may be confusing, but this notation is standard.) Show that for $A \subseteq X$, the sets $f(A^c)$ and $f(A)^c$ are not necessarily equal. In contrast, show that for any $B \subseteq Y$,

$$f^{-1}(B^c) \ = \ f^{-1}(B)^c.$$

Let $\{A_i\}$ be a nonempty family of subsets of X. Show that

$$f(\bigcup_i A_i) \ = \ \bigcup_i f(A_i) \quad \text{and} \quad f(\bigcap_i A_i) \ \subseteq \ \bigcap_i f(A_i).$$

Give an example where equality does not hold in the second case. Let $\{B_j\}$ be a nonempty family of subsets of Y. Show that

$$f^{-1}(\bigcup_j B_j) \;=\; \bigcup_j f^{-1}(B_j) \quad\text{and}\quad f^{-1}(\bigcap_j B_j) \;=\; \bigcap_j f^{-1}(B_j).$$

Observe that f^{-1} behaves better than f.

(2) Let $f : X \to Y$ and $g : Y \to Z$ and let $U \subseteq Z$. Show that

$$(g \circ f)^{-1}(U) \;=\; f^{-1}(g^{-1}(U)).$$

(3) Let $A_1, A_2, A_3, \ldots$ be subsets of a set X. Show that

$$\{x \mid x \in A_n \text{ for infinitely many } n\} \;=\; \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

(4) Show that the composite of two injective functions is injective. Same for surjective. (This yields two subcategories of Set.) Show that any function $h : X \to Z$ can be canonically factored as a composite of a surjective map followed by an injective map. That is there exists a commutative diagram

$$\begin{tikzcd} & Y \arrow[dr, hook, "g"] & \\ X \arrow[ur, two heads, "f"] \arrow[rr, "h"'] & & Z. \end{tikzcd}$$

The set Y is called the *(co)image* of h.

(5) Given any function $f : X \to Y$, describe all its left inverses, that is, functions $g : X \to X$ such that $f \circ g = \mathrm{id}$.

(6) How many functions are there from a finite set X to a finite set Y? What happens if either or both of X and Y are empty?

(7) A set A is *countable* if there exists a bijection between A and $\mathbb{N}$. Show that any infinite subset of a countable set is countable. Show that any countable union of countable sets is countable. Show that the set of rational numbers $\mathbb{Q}$ is countable.

(8) A set is *uncountable* if it is not finite or countable. Give an example of an uncountable set.

(9) For any set X, its power set denoted 2^X consists of all subsets of X. Show that a set X and its power set 2^X can never be isomorphic in Set, that is, there cannot exist any bijection between X and 2^X.

(10) Which objects in Set are isomorphic to $\mathbb{N}$? Describe in words: What are all automorphisms from $\mathbb{N}$ to itself?

(11) Check that the subcategory of finite sets is isomorphism-closed in Set.

(12) Show that there is a bijection between 2^X and the set of all functions from X to $\{0, 1\}$. Here X is any set (could be empty, could be infinite).

(13) Show that: Let X and Y be sets (possibly infinite). If $f : X \to Y$ and $g : Y \to X$ are injective, then show that there is a bijection from X to Y. (This is the Schroeder-Bernstein theorem.)

(14) Consider the following equivalence relation on $\mathbb{N} \times \mathbb{N}$:

$$(m, n) \sim (m', n') \iff m + n' = n + m'.$$

Check that this is an equivalence relation. Do you see that $(\mathbb{N} \times \mathbb{N})_\sim$ is the same as $\mathbb{Z}$. (One can in fact define $\mathbb{Z}$ in this manner.) How would you define the operations of addition and multiplication on these equivalence classes?

(15) Fix a positive integer n. Define $x \sim y$ if $x - y$ is divisible by n. Check that this is an equivalence relation on $\mathbb{Z}$. Describe the equivalence classes and the quotient map $\mathbb{Z} \to \mathbb{Z}_\sim$.

(16) Suppose a relation R on a set X is symmetric and transitive. We claim that R is reflexive by the following argument. Suppose we are given any $x \in X$. Pick a y such that $x \sim y$. By symmetry, $y \sim x$, and by transitivity $x \sim x$. What is wrong with this argument?

(17) On any set X, we have the empty relation $R = \emptyset$. What properties does this relation satisfy? That is, is it reflexive, irreflexive, ...?

(18) For objects a and b in a category C, we say $a \sim b$ if a is isomorphic to b. Show that this defines an equivalence relation on the set of objects of C. Can you describe the equivalence classes when $\mathsf{C} = \mathsf{Set}$ and when $\mathsf{C} = \mathsf{Vec}_\Bbbk$?

(19) Can you associate a category to any reflexive and transitive relation R on a set X?

1.3. Category of posets

We discuss posets in more detail. Zorn's lemma, which is an equivalent formulation of the axiom of choice, is phrased in the language of posets.

1.3.1. Partially ordered sets. A poset is a set P equipped with a relation which is transitive, reflexive, antisymmetric. Such a relation is usually denoted $\leq$. We read $x \leq y$ as "x is less than y", or "y is greater than x". In this notation, the poset axioms are as follows. For all $x, y, z \in P$,

- $x \leq x$,
- $x \leq y$ and $y \leq x$ implies $x = y$,
- $x \leq y$ and $y \leq z$ implies $x \leq z$.

Poset is a shortform for partially ordered set. This term was coined by Birkhoff in the third edition of his famous book on Lattice Theory (1940). If $x \leq y$ but $x \neq y$, then it is customary to write $x < y$.

Suppose P and Q are posets. A function $f : P \to Q$ is *order-preserving* if $x \leq y$ implies $f(x) \leq f(y)$. This defines the category Poset whose objects are posets, and morphisms are order-preserving maps. What do you need to check here?

What is the relationship between Poset and Set? Every poset is a set, and every order-preserving map is a function. Thus, there is a functor

$$\mathsf{Poset} \to \mathsf{Set}.$$

This is called a forgetful functor since it is constructed by forgetting the order relation of the poset.

1.3.2. Linearly ordered sets. A poset P is *linearly ordered* or *totally ordered* if for any $x, y \in P$ either $x \leq y$ or $y \leq x$. (This is a property of a poset as opposed to imposing more structure.) We can consider the full subcategory Lin of Poset whose objects are linearly ordered sets, and whose morphisms are order-preserving maps. Every subcategory gives rise to an inclusion functor. In this case, the functor is from the category of linearly ordered sets to the category of partially ordered sets.

The subcategory Lin is isomorphism-closed: If P is linearly ordered and $P \to Q$ is a poset isomorphism, then Q is necessarily linearly ordered.

1.3.3. Meets and joins. Let P be a poset.

- We say x is the *bottom element* of P if $x \leq y$ for all $y \in P$.
- Dually, we say x is the *top element* of P if $y \leq x$ for all $y \in P$.
- Given elements $x, y \in P$, the *meet* of x and y, denoted $x \wedge y$ is the largest element of P smaller than both x and y. That is, $x \wedge y \leq x$, $x \wedge y \leq y$, and ($z \leq x$ and $z \leq y$ implies $z \leq x \wedge y$).
- Dually, the *join* of x and y, denoted $x \vee y$ is the smallest element of P larger than both x and y.

Meets, joins, bottom and top elements may not exist, but they are unique whenever they exist.

More generally, one can start with any subset S of P, and define the *least upper bound* of S to be smallest element which is larger than all elements of S, and the *greatest lower bound* of S to be the largest element which is smaller than all elements of S. How would you intepret this statement when S is empty? The least upper bound of S is also called the join of the elements of S (since when S has two elements, least upper bound coincides with the join of the two elements). while the greatest lower bound is also called the meet of the elements of S.

An *upper bound* for S is an element of P which is larger than all elements of S. In general, S will have many upper bounds. We can look at the set whose elements consist of all upper bounds of S. The least upper bound of S is the bottom element of this set. Similar remarks apply to lower bounds and greatest lower bounds.

We have already mentioned how to associate a graph with directed edges to any relation. The graph of a poset is called a *Hasse diagram*. It is customary not to draw the loops and the arrows implied by transitivity. Further, bigger elements are written vertically above the smaller elements, which makes the arrows on the edges redundant.

Example 1.3. The set of natural numbers $\mathbb{N}$ is a linearly ordered set. It has a bottom element, namely 0, but no top element. The meet of two numbers is the smaller of the two, while the join is the larger of the two. The set of integers $\mathbb{Z}$ is also a linearly ordered set. It has no bottom or top element, while meets and joins are as above. The set of rational numbers $\mathbb{Q}$ is also a linearly ordered set.

Example 1.4. Let X be any set and 2^X denote its power set. An element of 2^X is a subset of X. This is a poset under inclusion, that is, we say $S \leq T$ if S is a subset of T. This is also called the *Boolean poset* of X. The bottom element is the empty set, the top element is X, meet is given by intersection of sets, and join by union of sets. Least upper bounds (that is, arbitrary joins) and greatest lower bounds are described the same way.

The Hasse diagram of the Boolean poset of $\{a, b, c\}$ is shown below.

$$
\begin{tikzcd}
 & \{a,b,c\} \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\{a,b\} \arrow[d, dash] \arrow[dr, dash] & \{a,c\} \arrow[dl, dash] \arrow[dr, dash] & \{b,c\} \arrow[dl, dash] \arrow[d, dash] \\
\{a\} \arrow[dr, dash] & \{b\} \arrow[d, dash] & \{c\} \arrow[dl, dash] \\
 & \emptyset &
\end{tikzcd}
$$

A poset is called a *lattice* if meets and joins (of any two elements) exist. (This is a property of a poset.) For instance, the Boolean poset is a lattice. A poset is called a *complete lattice* if arbitrary meets and joins exist. There is a distinction between lattices and complete lattices only if the poset is infinite.

1.3.4. Opposite poset. Given a poset $(P, \leq)$, define another relation $\leq'$ on P by

$$x \leq' y \text{ if } y \leq x.$$

Check that this defines a partial order on P. This is called the opposite partial order and P equipped with this order is called the opposite (or dual) poset. Note that in the passage to the opposite poset, bottom and top elements switch positions and so do meet and join.

A poset is called *self-dual* if it is isomorphic to its dual. For instance, the Boolean poset is self-dual (under taking complements).

Problems.

(1) Define the partial order on $\mathbb{N}_+$: $a \leq b$ if a divides b. What is the top element, bottom element, meet, join?

(2) A function between posets $f : P \to Q$ is *order-reversing* if $x \leq y$ implies $f(y) \leq f(x)$. Consider the gadget whose objects are posets and morphisms are order-reversing maps. Does this yield a category under usual composition of functions? What happens if we consider the gadget whose objects are posets and morphisms are either order-preserving or order-reversing maps.

(3) Describe all order-preserving maps from $\mathbb{N}$ to $\mathbb{Z}$ (or to any linearly ordered set).

(4) Let $(P, \leq)$ be a preposet. Define a relation on P as follows.

$$x \sim y \text{ if } x \leq y \text{ and } y \leq x.$$

(Recall that the relation in a preposet is not assumed to be antisymmetric.) Check that $\sim$ is an equivalence relation. Show that $\leq$ induces a partial order on the set of equivalence classes $P_{\sim}$.

(5) Describe the initial object, terminal object, product, coproduct in Poset.

(6) Every set X is a poset with the discrete partial order: $x \leq y$ iff $x = y$. Check that this defines a functor from Set to Poset. How does this relate to the forgetful functor from Poset to Set?

Can you extend the construction of a power set to a functor

$$\mathsf{Set} \to \mathsf{Poset}, \qquad X \mapsto 2^X?$$

(7) Let R be a relation on X and S be a relation on Y. A map $f : X \to Y$ is relation-preserving if xRy implies $f(x)Sf(y)$ for all $x, y \in X$. Show that the composite of relation-preserving maps is again relation-preserving. This yields a category whose objects are relations (on different sets) and morphisms are relation-preserving maps. How does this category relate to Poset?

(8) Show that no rational number satisfies the equation $x^2 = 2$. Consider the linearly ordered set $S = \{x \in \mathbb{Q} \mid x^2 < 2\}$ (with linear order inherited from $\mathbb{Q}$). Show that S does not have a top element. That is, if $\alpha^2 < 2$ for some rational α, then there exists a rational $\beta > \alpha$ such that $\beta^2 < 2$. Further, show that S does not have a least upper bound in the poset $\mathbb{Q}$. Can you think of some other subsets which do not have a least upper bound?

(9) Suppose P is a lattice. For any $x, y, z \in P$ show that $(x \vee y) \vee z = x \vee (y \vee z)$. Further show that this element is the least upper bound of the subset $S = x, y, z$. Extend this result to any finite set S. Deduce that if P is a finite poset then P is a lattice iff P is a complete lattice. Can you give an example of a lattice with a top and bottom element which is not complete?

(10) To each poset P associate a category C_P whose objects are elements of P, and there is a unique morphism from x to y whenever $x \leq y$. Check that the initial object, terminal object, product, coproduct exists in C_P iff the bottom element, top element, meet, join exists in P (respectively). Do you see that an order-preserving map $P \to Q$ yields a functor $\mathsf{C}_P \to \mathsf{C}_Q$. Thus, in fact, we dealing here with a functor from Poset to the category of all categories.

1.4. Monoids, groups, rings, fields

We discuss some important categories which are usually taught under the name 'abstract algebra'. You can find more information in [3, 10] (though not from a categorical viewpoint).

1.4.1. Monoids. A *monoid* is a set X with a distinguished element $e \in X$ (called the identity element) and a binary operation

$$X \times X \to X, \qquad (a, b) \mapsto a \cdot b,$$

which is associative and unital, that is,

$$(a \cdot b) \cdot c = a \cdot (b \cdot c), \quad \text{and} \quad a \cdot e = e \cdot a = a,$$

for all $a, b, c \in X$.

A monoid homomorphism is a function $f : X \to Y$ between monoids such that $f(e) = e$ and $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in X$.

This yields a category Monoid whose objects are monoids, and morphisms are monoid homomorphisms. What is it that you need to check?

A monoid is called *commutative* if $a \cdot b = b \cdot a$ for all $a, b \in X$. This is a property of a monoid. Thus, we have the full subcategory of commutative monoids.

Example 1.5. The set $\mathbb{N}$ is a commutative monoid under addition (with 0 being the identity element). The set $\mathbb{N}_+$ is a commutative monoid under multiplication (with 1 being the identity element).

Let X be the set of all (finite) words in the alphabet $\{0, 1\}$, that is,

$$X = \{\ , 0, 1, 00, 01, 10, 11, 000, 001, \dots\}.$$

The first element is the empty word. Then X is monoid under concatenation of words (writing words next to each other).

$$a = 001, \; b = 10, \quad a \cdot b = 00110.$$

The empty word is the identity.

For any S, let X be the set of all functions from S to itself. Then X is a monoid under composition of functions, the identity element is the identity function.

1.4.2. Groups. Let X be a monoid. An element $a \in X$ is called *invertible* if there exists an element $b \in X$ such that $a \cdot b = b \cdot a = e$. If this happens, we say that b is the inverse of a.

A monoid X is called a *group* if all elements of X are invertible. This is a property of a monoid. Thus, we have the full subcategory of groups denoted Group. In a group, we denote the inverse of an element a by a^{-1}.

A monoid X is called an *abelian group* if it is commutative and also a group. This defines the category of abelian group, it is a full subcategory of groups, as well as a full subcategory of commutative monoids (and of course, a full subcategory of monoids). It is traditional to use the term abelian group as opposed to commutative group.

Example 1.6. $\mathbb{N}$ under addition is *not* a group. Same with $\mathbb{N}_+$ under multiplication. $\mathbb{Z}$ under addition is an abelian group. However $\mathbb{Z}_+$ under multiplication is *not* a group. $\mathbb{Q}$ under addition and $\mathbb{Q}_+$ under multiplication are both abelian groups.

For any set S, the set of all bijections from S to itself denoted $\mathrm{Aut}(S)$ is a group, with the binary operation given by composition of functions. This is called the automorphism group of the set S. This group is *not* abelian if S has more than two elements.

Perhaps add standard cancellation properties.

1.4.3. Rings. A *ring* is a set R with distinguished elements $0, 1 \in R$ and two binary operations

$$R \times R \to R, \qquad (a, b) \mapsto a + b, \quad \text{and} \quad (a, b) \mapsto a \cdot b$$

called addition and multiplication such that R is an abelian group under addition with 0 as the identity element, R is a monoid under multiplication with 1 as the identity element, and

$$a \cdot (b + c) = a \cdot b + a \cdot c, \quad \text{and} \quad (a + b) \cdot c = a \cdot c + b \cdot c,$$

for all $a, b, c \in R$. (These two identities are called the distributive law.)

A *ring homomorphism* is a function $f : R \to S$ between rings such that f is a monoid homomorphism with respect to both addition and multiplication.

This defines the category of rings denoted Ring.

A ring is called *commutative* if the underlying monoid under multiplication is commutative: $a \cdot b = b \cdot a$ for all $a, b \in R$. (Note that the underlying monoid under addition is always commutative.) Thus, we have the full subcategory of commutative rings. Commutative algebra/Algebraic geometry is the study of this category.

Example 1.7. $\mathbb{Z}$ under the usual operations of addition and multiplication is a commutative ring. The set of polynomials in one variable with integer coefficients, denoted $\mathbb{Z}[x]$, is a commutative ring under the usual operation of addition and multiplication of polynomials. Fix an integer α. Then the evaluation map

$$\mathbb{Z}[x] \to \mathbb{Z}, \qquad f(x) \mapsto f(\alpha)$$

is a ring homomorphism. What do you need to check here? In particular, we can take $\alpha = 0$. The map then sends a polynomial to its constant term.

$\mathbb{Q}$ is also a ring, and the usual inclusion map $\mathbb{Z} \to \mathbb{Q}$ is a ring homomorphism. Since this map is injective, we say that the addition and multiplication operations of $\mathbb{Q}$ extend the corresponding operations of $\mathbb{Z}$.

The set of all 2×2 matrices with integer coefficients is a (noncommutative) ring under the usual operations of matrix addition and multiplication. Can you think of other examples of noncommutative rings?

Note that $\mathbb{N}$ has two operations but it is not a ring because it is not a group under addition. It is what is called a *rig*. A rig only requires a commutative monoid under addition. Mathematicians apparently forgot to study this notion.

1.4.4. Fields. A commutative ring is a *field* if $0 \neq 1$ and $R \setminus 0$ is an abelian group under multiplication with 1 being the identity element. In other words, a field is a commutative ring in which $0 \neq 1$ and every nonzero element has a multiplicative inverse. We may also say: A commutative ring is a field if $0 \neq 1$ and every nonzero element has a multiplicative inverse. This defines the full subcategory Field.

Example 1.8. The commutative ring $\mathbb{Z}$ is *not* a field, however the commutative ring $\mathbb{Q}$ is a field.

The polynomial ring $\mathbb{Z}[x]$ is *not* a field. We do not get a field even if we allow the coefficients to be rational numbers. That is, $\mathbb{Q}[x]$ is also *not* a field. We get a field if we consider rational functions: elements are (equivalence classes of) fractions $\frac{p(x)}{q(x)}$, where $p(x), q(x) \in \mathbb{Z}[x]$ and $q(x) \neq 0$. The inverse of $\frac{p(x)}{q(x)}$ is $\frac{q(x)}{p(x)}$. This is an example of localization mentioned earlier.

Suppose E and F are fields and $f : E \to F$ is a ring homomorphism (that is, a morphism in Field). We claim that $f(a) = 0$ implies $a = 0$. Suppose $f(a) = 0$ but $a \neq 0$. Let b be the multiplicative inverse of a. Then $1 = f(ab) = f(a)f(b) = 0f(b) = 0$. This is contradiction, so the claim holds. We further claim that f is injective: Suppose $f(b) = f(c)$. Then $f(b-c) = 0$, and hence by the claim $b-c = 0$. Thus a morphism between fields is necessarily injective, and it makes sense to say that F is a larger field than E. The term that is commonly used is "field extension".

The category of fields has no initial or final objects.

Elements of a field are often called scalars. In physics, the term scalar field on a space X simply means a function from X to some field.

1.4.5. Vector spaces. Fix a field $\Bbbk$. A *vector space* over $\Bbbk$ is an abelian group (whose binary operation is denoted $(v, w) \mapsto v + w$) equipped with an operation

$$\Bbbk \times V \to V, \qquad (a, v) \mapsto av$$

called scalar multiplication such that

$$1v = v, \quad (ab)v = a(bv), \qquad a(v + w) = av + aw, \quad (a + b)v = av + bv$$

for all $a, b \in \Bbbk$ and $v, w \in V$. A $\Bbbk$-*linear map* between V and W is a function $f : V \to W$ such that

$$af(v) = f(av), \qquad f(v + w) = f(v) + f(w)$$

for all $a \in \Bbbk$ and $v, w \in V$. Equivalently, a $\Bbbk$-linear map f is a group homomorphism such that $af(v) = f(av)$. This defines the category of vector spaces denoted $\mathsf{Vec}_{\Bbbk}$: objects are vector spaces over $\Bbbk$ and morphisms are $\Bbbk$-linear maps.

Observe that the above definitions can be made for any ring $\Bbbk$. In the general context of rings, by tradition, the term "module" is used instead of "vector space", and "module homomorphism" instead of "linear map". The category of modules is usually denoted $\mathsf{Mod}_{\Bbbk}$.

1.4.6. Ordered fields. An *ordered field* is a field $\Bbbk$ equipped with a linear order $\leq$ such that for any $x, y, z \in \Bbbk$

$$y < z \text{ implies } x + y < x + z \quad \text{and} \quad x > 0, y > 0 \text{ implies } xy > 0. \tag{1.2}$$

A morphism between ordered fields is a ring homomorphism which is order-preserving. This defines the category of ordered fields. $\mathbb{Q}$ is an example of an ordered field.

An element x is called *positive* if $x > 0$ and *negative* if $x < 0$. Familiar ways of manipulating inequalities work in any ordered field. For instance, multiplication by a positive element preserves an inequality, while multiplication by a negative element reverses an inequality, For more, see [21, Proposition 1.18].

Problems.

(1) In a monoid X, show that any invertible element has a unique inverse.
(2) Suppose $f : G \to H$ is a function between groups such that $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in G$. Show that $f(e) = e$ (so f is a group homomorphism), and $f(a^{-1}) = f(a)^{-1}$ for all $a \in G$. Show that the first assertion may fail for monoids.
(3) Show that a monoid homomorphism $f : X \to Y$ is an isomorphism iff f is a bijection.
(4) For a monoid M, let $M^{\times}$ denote the set of invertible elements of M. Show that $M^{\times}$ is a group with the binary operation induced from M. Moreover, a monoid homomorphism $M \to N$ induces a group homomorphism $M^{\times} \to N^{\times}$. In other words, there is a functor $\mathsf{Monoid} \to \mathsf{Group}$. How does this relate to the inclusion functor $\mathsf{Group} \to \mathsf{Monoid}$?
(5) Let S be a set. Think of it as an alphabet. Let $\mathcal{F}(S)$ be the set consisting of all words written using the alphabet S. Check that the operation of concatenation of words (writing one word after the other) turns S into a monoid. What is the identity element? Which familiar monoid do you get when S is a singleton set? Now suppose S and T are two sets and $f : S \to T$ is any function. This induces a function denoted $\mathcal{F}(f)$ from $\mathcal{F}(S)$ to $\mathcal{F}(T)$. Check that $\mathcal{F}$ defines a functor from Set to Monoid. We also have the forgetful functor in the other direction from Monoid to Set. How do the two functors relate?
(6) Show that in any ring R, $0 \cdot a = 0 = a \cdot 0$ for all $a \in R$. Deduce that if $0 = 1$ in a ring R, then R is a singleton $\{0\}$. (This is called the *zero ring*.)
(7) To every ring R, let $R[x]$ denote the set of polynomials with coefficients in R. Define the operations of addition and multiplication in $R[x]$ similar to polynomials with integer coefficients. (Note how the R instead of those in $\mathbb{Z}$ get used.) Show that $R[x]$ is a ring. Further, show that we obtain a functor

$$\mathsf{Ring} \to \mathsf{Ring}, \qquad R \mapsto R[x].$$

(8) Describe the initial object, terminal object, product, coproduct in the categories $\mathsf{Vec}_{\Bbbk}$, Group, commutative Ring.

(9) Construct a field with p elements, where p is any prime number.

(10) What are all ring homomorphisms from $\mathbb{Q}$ to itself?

(11) Fix a field $\Bbbk$. Let $\Bbbk^n$ denote the n-fold cartesian product of $\Bbbk$, that is, elements of $\Bbbk^n$ are n-tuples with entries in $\Bbbk$. Check that $\Bbbk^n$ is a vector space over $\Bbbk$ with component-wise addition and scalar multiplication. (Here $n = 0, 1, 2, \ldots$, $\Bbbk^0$ is the zero vector space $\{0\}$.) Show that there is a canonical bijection between the set of $\Bbbk$-linear maps from $\Bbbk^n$ to $\Bbbk^m$, and the set of $m \times n$ matrices with entries in $\Bbbk$. Further show that composition of $\Bbbk$-linear maps between such vector spaces corresponds to matrix multiplication.

(12) Fix a field $\Bbbk$. To each set X, one can define a vector space, denoted $\Bbbk X$, whose elements are formal finite linear combinations

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n,$$

where $a_i \in \Bbbk$ and $x_i \in X$. This is the same as assigning a scalar to each $x \in X$ such that only finitely many of these scalars are nonzero. How would you define addition and scalar multiplication? Note that X could be infinite. Extend this construction to a functor

$$\mathsf{Set} \to \mathsf{Vec}_{\Bbbk}, \qquad X \mapsto \Bbbk X.$$

How does this relate to the forgetful functor from $\mathsf{Vec}_{\Bbbk}$ to Set?

(13) To each vector space V over a field $\Bbbk$, one can associate a poset whose elements are subspaces of V partially ordered by set-inclusion. (A *subspace* of V is a subset which is closed under addition and scalar multiplication.) Is this poset a lattice? Does it have a top element and a bottom element? Can you extend this construction to a functor

$$\mathsf{Vec}_{\Bbbk} \to \mathsf{Poset}?$$

Similar constructions can be made in other contexts, for instance, to a group one may associate the poset of subgroups.

(14) Fix a field $\Bbbk$. A field extension of $\Bbbk$ is also a $\Bbbk$-vector space (with scalar multiplication defined using the multiplication in the field.) In particular, every field is a vector space over itself. Let $\mathsf{Field}_{\Bbbk}$ denote the subcategory of Field consisting of all field extensions of $\Bbbk$, and those ring homomorphisms which fix $\Bbbk$. Show that this yields a functor

$$\mathsf{Field}_{\Bbbk} \to \mathsf{Vec}_{\Bbbk}.$$

CHAPTER 2

Real numbers

We construct the real number system. It has the structure of an ordered field which satisfies the least upper bound property. We also discuss binary expansions. expand later

2.1. Convergent and Cauchy sequences

We begin with sequences in the set of rational numbers, and the notions of convergence and Cauchyness therein.

2.1.1. Sequences. Let X be any set. A sequence in X is a function $a : \mathbb{N} \to X$. It is usually denoted $(a_0, a_1, a_2, \dots)$, with $a_n := a(n)$, the value of the function a at n. As a shorthand, we denote a sequence by (a_n). Traditionally, a sequence is defined as a function $a : \mathbb{N}_+ \to X$, so that it reads $(a_1, a_2, \dots)$. However, since

$$\mathbb{N} \to \mathbb{N}_+, \qquad n \mapsto n+1 \tag{2.1}$$

is a bijection, $\mathbb{N}$ and $\mathbb{N}_+$ are isomorphic in the category of sets. So one can pass from one definition to the other via this bijection. We will use either definition depending on our convenience. (Note that to define a sequence, one could use any countable set instead of $\mathbb{N}$.)

A *subsequence* of a sequence (a_n) is a sequence of the form (a_{i_n}) for some $0 \le i_1 < i_2 < \dots$. Here $i_1, i_2, \dots$ are nonnegative integers. This definition uses the linear order on $\mathbb{N}$. Note in this regard that (2.1) is an isomorphism of posets.

2.1.2. Convergence of sequences in $\mathbb{Q}$. A sequence (a_n) in $\mathbb{Q}$, that is, $a_n \in \mathbb{Q}$ for all n, is *convergent* if there exists a rational number a such that for every rational number $\epsilon > 0$, there exists an integer N such that $|a_n - a| < \epsilon$ for all $n > N$. The rational number a is called the *limit* of the sequence (a_n), and we say that (a_n) converges to a. It is convenient to denote this by $a_n \to a$. Another standard notation is

$$\lim_{n\to\infty} a_n = a.$$

Lemma 2.1. *The limit of a sequence, if it exists, is unique.*

Proof. Suppose $a_n \to a$ and $a_n \to b$ and $a \neq b$. Then take a rational number $\epsilon > 0$ smaller than half of $b - a$ and get a contradiction. $\square$

Lemma 2.2. *If $a_n \to a$ and $b_n \to b$, then $a_n + b_n \to a + b$ and $a_n b_n \to ab$. If $a_n \to a$, and $a \neq 0$ and $a_n \neq 0$ for all n, then $(1/a_n)$ converges to $1/a$.*

Proof. For the first statement, use the estimates

$$|(a+b) - (a_n + b_n)| = |(a - a_n) + (b - b_n)| \le |a - a_n| + |b - b_n|$$

and

$$|ab - a_n b_n| = |(a - a_n)b + a_n(b - b_n)| \leq |(a - a_n)b| + |a_n(b - b_n)|.$$

Details and the second statement are left as an exercise. □

The notation $a_n \not\to a$ means that (a_n) does not converge to a. (In this case, the sequence is either not convergent or converges to some other point.) The statement $a_n \not\to a$ is equivalent to: There exists a rational number $\epsilon > 0$ and a subsequence (a_{i_n}) such that $|a_{i_n} - a| > \epsilon$ for all n.

2.1.3. Cauchy sequences in $\mathbb{Q}$. A sequence (a_n) in $\mathbb{Q}$ is *Cauchy* if for every rational number $\epsilon > 0$, there exists an integer N such that $|a_m - a_n| < \epsilon$ for all $m, n > N$.

Lemma 2.3. *If (a_n) and (b_n) are Cauchy, then so are $(a_n + b_n)$ and $(a_n b_n)$. If (a_n) is Cauchy and $a_n \neq 0$ for all n and $a_n \not\to 0$, then $(1/a_n)$ is Cauchy (and does not converge to 0).*

Proof. Similar to that of Lemma 2.2. □

Intuitively, a Cauchy sequence is a sequence all of whose terms come close to one another. Note that Cauchyness is defined purely using elements of the sequence and makes no reference to any limit.

Lemma 2.4. *If a sequence in $\mathbb{Q}$ is convergent, then it is Cauchy.*

Proof. Suppose (a_n) is a sequence which converges to a. Let $\epsilon > 0$ be given. Choose N such that for all $n > N$, $|a_n - a| < \epsilon/2$. Now for any $m, n > N$,

$$|a_m - a_n| \leq |a_m - a| + |a_n - a| < \epsilon/2 + \epsilon/2 = \epsilon.$$

So (a_n) is Cauchy. □

However a Cauchy sequence in $\mathbb{Q}$ need not be convergent. Consider the sequence defined recursively by

$$a_0 = 1, \quad a_{n+1} = \frac{1}{2}(a_n + \frac{2}{a_n}) \text{ for all } n \geq 1.$$

Check that this sequence is decreasing from the first term, that is, $a_1 \geq a_2 \geq \dots$ and bounded below by 0. Hence it is Cauchy. Why? Check that (a_n^2) converges to 2. It follows that (a_n) does not converge in $\mathbb{Q}$ (because if it did converge to say a, then by Lemma 2.2, $a^2 = 2$ yielding a contradiction). Can you think of other examples of this kind?

2.1.4. Bounded sequences. A sequence (a_n) in $\mathbb{Q}$ is *bounded* if it has an upper bound and a lower bound in the poset $(\mathbb{Q}, \leq)$. Explicitly, there exist rational numbers M_1 and M_2 such that

$$M_1 \leq a_n \leq M_2$$

for all n.

Lemma 2.5. *Any Cauchy sequence (and hence any convergent sequence) is bounded.*

Proof. Let (a_n) be a Cauchy sequence. Apply the Cauchy condition to say $\epsilon = 1$. Thus there exists N_0 be such that $|a_m - a_n| < 1$ for all $m, n > N_0$. Let M be any rational less than the finitely many terms upto a_{N_0} and also $a_{N_0} - 5$ (say). Then M is a lower bound for (a_n). (We can afford to be sloppy about our choices here.) The existence of the upper bound is similar. □

2.2. Real numbers

What are the real numbers? This is very vague. More precise would be: Is there a set whose elements are the real numbers? The answer is yes. Let us denote this set by $\mathbb{R}$. Next question: What extra structure does $\mathbb{R}$ possess? Is it a monoid? Is it a field? and so on. The answer is that $\mathbb{R}$ is an object in all the categories listed in Table 1.1. Thus, it is a monoid, a group, ..., and a topological space.

There are two well-known approaches for constructing the real numbers: one using Cauchy sequences due to Cantor, and another using Dedekind cuts due to Dedekind. Both constructions appeared around the same time (1872). We discuss the approach of Cantor.

2.2.1. Set of real numbers. Let Ω denote the set of all Cauchy sequences in $\mathbb{Q}$. Define an equivalence relation on Ω as follows.

$$(a_n) \sim (b_n) \;\text{ if }\; \lim_{n\to\infty} a_n - b_n = 0.$$

The condition means that $(a_n - b_n)$ converges to 0, that is, given any rational number $\epsilon > 0$, there exists an integer N such that $|a_n - b_n| < \epsilon$ for all $n > N$. By Lemma 2.2, we deduce that this is an equivalence relation.

Observation 2.6. If (a_n) converges to a, then every subsequence of (a_n) also converges to a. Similarly, if (a_n) is a Cauchy sequence, then every subsequence of (a_n) is also a Cauchy sequence and $\sim$-equivalent to (a_n).

Define the set of real numbers, denoted $\mathbb{R}$, to be the set of equivalence classes $\Omega_\sim$. In other words, a real number is an equivalence class of Cauchy sequences in $\mathbb{Q}$. We say that a Cauchy sequence (a_n) represents a real number r if $(a_n) \in r$.

Every rational number a gives rise to the constant Cauchy sequence $(a, a, a, \dots)$. This yields a map

$$\mathbb{Q} \hookrightarrow \mathbb{R}, \qquad a \mapsto [(a, a, a, \dots)]. \tag{2.2}$$

The square brackets around the sequence are used to indicate the equivalence class of the sequence. This map is injective. Why? A real number is called *irrational* if it is not in the image of this map.

2.2.2. Field operations. Define the operations of addition and multiplication on $\mathbb{R}$ by

$$[(a_n)] + [(b_n)] := [(a_n + b_n)] \qquad \text{and} \qquad [(a_n)][(b_n)] := [(a_n b_n)].$$

In view of Observation 1.1, we need to check the following.

- If (a_n) and (b_n) are Cauchy, then so are $(a_n + b_n)$ and $(a_n b_n)$.
- If $(a_n) \sim (a'_n)$ and $(b_n) \sim (b'_n)$, then $(a_n + b_n) \sim (a'_n + b'_n)$ and $(a_n b_n) \sim (a'_n b'_n)$.

These follow from Lemmas 2.2 and 2.3.

Proposition 2.7. *The set of real numbers $\mathbb{R}$ is a field (with addition and multiplication as defined above). Further, (2.2) is a ring homomorphism.*

The second statement is usually stated as: $\mathbb{Q}$ is a subfield of $\mathbb{R}$ via (2.2). (Whenever we have an injective map $f : X \to Y$, $f(X)$ is a subset of Y, but we abuse precision and say X is a subset of Y. This is alright since X and $f(X)$ are bijective.)

PROOF. Observe that $\mathbb{R}$ under addition is an abelian group. The identity element is the equivalence class of the constant sequence $(0, 0, \dots)$, and the inverse of $[(a_n)]$ is $[(-a_n)]$. Similarly, $\mathbb{R}$ is a commutative monoid under multiplication, with identity element being the equivalence class of the constant sequence $(1, 1, \dots)$. The distributive law clearly holds. So $\mathbb{R}$ is a ring. To show that it is a field, we need to check that every nonzero element of $\mathbb{R}$ is invertible. Accordingly, let (a_n) be a Cauchy sequence which does not converge to zero. Then only finitely many elements of this sequence will be zero. (If not, then it will have a subsequence converging to zero contradicting Observation 2.6.) Throwing them out, we get a subsequence (a'_n) which is equivalent to (a_n). Then $(\frac{1}{a'_n})$ is also a Cauchy sequence (which does not converge to zero), and is the required inverse. □

2.2.3. Linear order on $\mathbb{R}$. We say that a real number r is *positive* if there exists a rational number $\epsilon > 0$ and a representative (a_n) of r such that $a_n > \epsilon$ for all n. Observe that:

(1) The sum of two positive numbers is positive.
(2) 0 is not positive.
(3) A real number r and its additive inverse $-r$ cannot both be positive. (This follows from the previous two facts.)

Given real numbers r and s, we say that $r \leq s$ if $s - r$ is either positive or zero.

Proposition 2.8. *$(\mathbb{R}, \leq)$ is a linearly ordered set.*

PROOF. We first verify that $\leq$ is a partial order.

- Reflexivity. Clearly $r \leq r$.
- Transitivity. We want: $r \leq s$ and $s \leq t$ implies $r \leq t$. This is clear if either $r = s$ or $s = t$. So suppose that $s - r$ and $t - s$ are both positive. Then by (1) above, their sum $t - r$ is positive. Hence $r \leq t$ as required.
- Antisymmetry. Suppose $r \leq s$ and $s \leq r$. We want to show that $s = r$. Suppose not. Then $s - r$ and $r - s$ are both positive. But this contradicts (3) above.

Conclusion: $\leq$ is a partial order.

Now we show that $\leq$ is a linear order, that is, for all real numbers r and s, either $r \leq s$ or $s \leq r$. Suppose $r \not\leq s$, that is, $r \neq s$ and $s - r$ is not positive. Pick any representative (a_n) of $s - r$. We claim that there is a $\epsilon > 0$ and a subsequence of (a_n) which lies entirely below $-\epsilon$. Two very similar arguments are given below.

- Since $r \neq s$, (a_n) cannot converge to 0. So there exists a $\epsilon > 0$ for which no N works in the convergence definition. This yields a subsequence (b_n) of (a_n) whose elements lie outside of $(-\epsilon, \epsilon)$. This subsequence also represents $s - r$. Since $s - r$ is not positive, only finitely many elements of (b_n) can be greater than ϵ (else we will get a subsequence entirely above ϵ). By throwing these out, we get a further subsequence which lies entirely below $-\epsilon$.
- Since $s - r$ is not positive, for any rational $\epsilon > 0$, some a_i will be smaller than ϵ. Applying this to $\epsilon = 1, 1/2, 1/3, \dots$, pick a subsequence (a_{i_n}) such that $a_{i_n} < 1/n$ for all positive integers. Since $r \neq s$, it cannot converge to 0. This yields a further subsequence of (a_{i_n}) whose elements lie outside of $(-\epsilon, \epsilon)$. Combining this with $a_{i_n} < 1/n$, note that eventually this subsequence must lie entirely below $-\epsilon$.

This proves the claim. Since the subsequence given by the claim also represents $s - r$, we conclude by taking additive inverses, that $r - s$ is positive. $\square$

Proposition 2.9. $(\mathbb{R}, \leq, +, \times)$ *is an ordered field.*

Proof. We need to check (1.2), but both properties are evident. $\square$

Note that $r < s$ iff $s - r$ is positive (since $r - s = 0$ iff $r = s$). Thus our usage of the term "positive" is consistent with the usage of "positive" is an ordered field. Continuing, we say that a real number r is *negative* if $r < 0$. The linear order on $\mathbb{R}$ implies that any real number is either positive, negative or zero.

The poset $(\mathbb{R}, \leq)$ is self-dual. For instance, the map

$$\mathbb{R} \to \mathbb{R}, \qquad x \mapsto -x$$

is an order-reversing isomorphism. So any statement about (least) upper bounds in $\mathbb{R}$ is equivalent to a statement about (greatest) lower bounds.

Lemma 2.10. *For real numbers r and s, $r \leq s$ iff there exists a representative (a_n) of r and a representative (b_n) of s such that $a_n \leq b_n$ for all n.*

The weaker looking condition $a_n \leq b_n$ for all $n > 20$ (say) is equivalent to the above (since we can throw away the first twenty terms of either sequence without changing their equivalence classes).

Proof. Forward implication. If $r = s$, then the assertion is clear. So suppose $s - r$ is positive. Then there exists a sequence (c_n) representing $s - r$ such that $c_n > 0$ for all n. Pick any representative (a_n) of r. Put $b_n := a_n + c_n$. Then (b_n) represents s and $a_n \leq b_n$ for all n.

Backward implication. The sequence $(b_n - a_n)$ consists of nonnegative rational numbers. If it converges to 0, then $r = s$ which is fine. Suppose not. Then there exists a rational $\epsilon > 0$ and a subsequence $(b_{i_n} - a_{i_n})$ such that $b_{i_n} - a_{i_n} > 0$ for all n. Since this subsequence represents $s - r$, we conclude that $r < s$. $\square$

We define the *absolute value* of a real number r, denoted $|r|$, by

$$|r| := \begin{cases} r & \text{if } r \text{ is positive,} \\ -r & \text{if } r \text{ is negative,} \\ 0 & \text{if } r \text{ is zero.} \end{cases} \tag{2.3}$$

This is a function from $\mathbb{R}$ to itself. The following properties are easy to verify.

$$\begin{gathered} |r| \geq 0, \text{ and it is zero iff } r = 0. \\ |r| < s \text{ iff } -s < |r| < s. \\ |r + s| \leq |r| + |s|. \end{gathered} \tag{2.4}$$

Lemma 2.11. *Let r be a real number, and (a_n) be a Cauchy sequence. Then (a_n) represents r iff for any rational number $\epsilon > 0$, there exists a N such that $|r - a_n| < \epsilon$ for all $n > N$.*

Proof. Forward implication. Let $\epsilon > 0$ be given. Use Cauchyness to pick N such that $-\epsilon/2 < a_m - a_n < \epsilon/2$. For each n, the sequence $(a_m - a_n)$ (with variable being m and n fixed) represents $r - a_n$, so applying Lemma 2.10, we deduce that $-\epsilon/2 \leq r - a_n \leq \epsilon/2$ for all $n > N$. This yields $|r - a_n| < \epsilon$ as required. Why did we use $\epsilon/2$ instead of ϵ in the argument?

Backward implication. Pick a representative (b_n) of r. Then by triangle inequality above, $|a_n - b_n| \leq |r - a_n| + |r - b_n|$. We can control $|r - a_n|$ by hypothesis, and $|r - b_n|$ because of the forward implication. Hence $a_n - b_n \to 0$, and (a_n) also represents r.

$\square$

Lemma 2.12. *Suppose (a_n) is a Cauchy sequence representing r. If s is a real number such that $a_n \leq s$ for all n, then $r \leq s$. Dually: If s is a real number such that $s \leq a_n$ for all n, then $s \leq r$.*

Proof. Pick a representative (b_n) of s. Define $b'_n := \max\{a_n, b_n\}$. Then we claim that (b'_n) also represents s. For this, we need to show that $b'_n - b_n \to 0$. Let a rational $\epsilon > 0$ be given. The term $b'_n - b_n$ is $a_n - b_n$ when $a_n \geq b_n$, and 0 otherwise. In the former case, $0 \leq a_n - b_n \leq s - b_n$. By Lemma 2.11, there exists a N such that $|s - b_n| < \epsilon$ for $n > N$. Hence $|b'_n - b_n| < \epsilon$ for $n > N$. So the claim follows. Now $a_n \leq b'_n$ for all n, so by Lemma 2.10, we have $r \leq s$.

The second statement follows from the first from self-duality of $\mathbb{R}$. $\square$

2.2.4. Binary representation of real numbers. Fix a real number r. We now give an algorithm whose output is an infinite string involving 0 and 1.

- Let a_0 be the largest integer such that $a_0 \leq r$. (Use Lemmas 2.5 and 2.12.)
- If $a_0 + \frac{1}{2} \leq r$, then put $a_1 := a_0 + \frac{1}{2}$ and $d_1 := 1$, else put $a_1 := a_0$ and $d_1 := 0$.
- If $a_1 + \frac{1}{4} \leq r$, then put $a_2 := a_1 + \frac{1}{4}$ and $d_2 := 1$, else put $a_2 := a_1$ and $d_2 := 0$.

We continue in this manner to get a sequence $(a_0, a_1, a_2, \dots)$ in $\mathbb{Q}$ and a sequence $(d_1, d_2, \dots)$ in $\{0, 1\}$.

Lemma 2.13. *Given a real number r, the sequence (a_n) as constructed above is Cauchy and represents r.*

This construction thus picks a canonical representative for each real number. Compare this with the situation for rational numbers, where the canonical representatives are fractions in lowest terms.

Proof. Using the estimate

$$\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^i} < 1 \qquad \text{for any } i \geq 1,$$

one can show that (a_n) is Cauchy. Let (b_n) be the sequence defined by

$$b_n := a_n + \frac{1}{2^n}. \tag{2.5}$$

Since (b_n) is a sum of two Cauchy sequences, it is also Cauchy by Lemma 2.3. Note that (a_n) and (b_n) represent the same real number. Call it s. Further, by construction,

$$a_n \leq r < b_n.$$

So by Lemma 2.12, $s \leq r \leq s$. So by antisymmetry, $r = s$, as required. (We are trapping r between two Cauchy sequences using repeated bisection.)

Alternatively, without introducing the b_n, we can directly note that $r - a_n < \frac{1}{2^n}$, and use the backward implication of Lemma 2.11. $\square$

Observe that for $n \geq 1$,

$$a_n = a_0 + \frac{d_1}{2} + \cdots + \frac{d_n}{2^n}. \tag{2.6}$$

Thus the d_i uniquely determine a_n and hence r. We use the notation

$$a_0 \cdot d_1 d_2 \dots$$

for the Cauchy sequence (a_n). This is called the binary expansion of r, a_0 is called the integer part, and d_i are the digits in the binary expansion. We could also write a_0 in binary using nonnegative powers of 2, but here our focus is on what happens after a_0.

Lemma 2.14. *Distinct real numbers have distinct binary expansions.*

Proof. Suppose $r \neq s$. Without loss of generality, assume that $r < s$. Then there exists a j such that $r < r + \frac{1}{2^j} < s$. Check that the binary expansions of r and s necessarily differ either in the integer part, or somewhere in the first j digits.

Alternatively, we could argue more abstractly as follows. Suppose r and s have the same binary expansion, that is, they produce the same sequence (a_n). Then by Lemma 2.13, (a_n) represents both r and s. Hence $r = s$. $\square$

Lemma 2.15. *Suppose $a_0 \cdot d_1 d_2 \dots$ is given, where a_0 is any integer and each d_i is either 0 or 1. Then it is the binary expansion of a real number iff the d_i do not end in any infinite string of 1's.*

Proof. Define the sequence (a_n) using (2.6). Then (a_n) is Cauchy, so it represents a real number, say r. However, the binary expansion of r will agree with the given $a_0 \cdot d_1 d_2 \dots$ iff the d_i do not end in any infinite string of 1's. We leave the details as an exercise. $\square$

The set $\mathbb{R}$ is in bijection with the set of symbols of the form $a_0 \cdot d_1 d_2 \dots$, where a_0 is any integer, each d_i is either 0 or 1, and the d_i do not end in any infinite string of 1's.

Instead of bisections, we could formulate the algorithm using trisections. Powers of 3 now come in, and the digits d_i take values in $\{0, 1, 2\}$. This is called a ternary expansion. If we use powers of 10, then we get the familiar decimal expansions.

2.2.5. Least upper bound property. The following is a key order-theoretic property of $\mathbb{R}$.

Proposition 2.16. *Any nonempty subset of $\mathbb{R}$ which is bounded above has a least upper bound. Dually, any nonempty subset of $\mathbb{R}$ which is bounded below has a greatest lower bound.*

The least upper bound is also called the *supremum*, and the greatest lower bound is called the *infimum*.

Proof. We prove the dual statement. Suppose S is the given nonempty subset. We follow a method which is very similar to the algorithm for the binary expansion.

- Let a_0 be the largest integer such that a_0 is a lower bound for S. This exists because S has a lower bound by hypothesis.
- If $a_0 + \frac{1}{2}$ is a lower bound for S, then put $a_1 := a_0 + \frac{1}{2}$, else put $a_1 := a_0$.
- If $a_1 + \frac{1}{4}$ is a lower bound for S, then put $a_2 := a_1 + \frac{1}{4}$, else put $a_2 := a_1$.

We continue in this manner to get a Cauchy sequence (a_n). Let r be its equivalence class. We claim that r is the greatest lower bound for S. Firstly, since each a_n is a lower bound for S, so is r by Lemma 2.12. Define b_n using (2.5). Then (b_n) also represents r but no b_n is a lower bound for S by construction. So r must be the greatest lower bound: If not, then let s be lower bound for S and $r < s$. Pick $\epsilon > 0$ such that $r + \epsilon < s$. By Lemma 2.11, the b_n eventually will be smaller than $r + \epsilon$, and this is a contradiction. $\square$

Note that $\mathbb{R}$ as a poset does not have a top element or a bottom element. So let us enlarge $\mathbb{R}$ to a bigger poset $\hat{\mathbb{R}}$ by adding two elements denoted $+\infty$ and $-\infty$, where $+\infty$ is the top element, and $-\infty$ is the bottom element. It is called the *extended real number system.*

Proposition 2.17. *The extended real number system $\hat{\mathbb{R}}$ is a complete lattice.*

Proof. This is more or less a reformulation of Proposition 2.16. Let S be any subset of $\hat{\mathbb{R}}$. We show that S has a least upper bound in $\hat{\mathbb{R}}$. If $\infty \in S$, then the least upper bound is ∞. If S is the empty set or the singleton $\{-\infty\}$, then the least upper bound is $-\infty$. So assume that S does not contain ∞ and contains at least one real number. We may further assume that S does not contain $-\infty$. (If it did, then we can remove it without affecting the set of upper bounds of S.) In other words, we assume that S is a nonempty subset of $\mathbb{R}$. If S does not have an upper bound in $\mathbb{R}$, then the least upper bound of S is ∞. If S does have an upper bound in $\mathbb{R}$, then we apply Proposition 2.16 to conclude that S has a least upper bound in $\mathbb{R}$ and hence in $\hat{\mathbb{R}}$. $\square$

Given any poset, there is a "smallest" complete lattice which contains the given poset. This is called the Dedekind–MacNeille completion of the poset. The Dedekind–MacNeille completion of the poset $\mathbb{Q}$ is precisely $\hat{\mathbb{R}}$. This is Dedekind's approach to the construction of the real numbers.

2.2.6. Existence of nth roots.

Proposition 2.18. *Fix a positive integer n. For any positive real number s, there is a unique positive real number r such that $r^n = s$.*

This is [21, Theorem 1.21]. The proof there is derived as a formal consequence of the least upper bound property. The proof below is algorithmic.

Proof. For simplicity, let us take $n = 2$. Uniqueness is clear since $0 < r < t$ implies $r^2 < t^2$. To prove existence, we start constructing the binary expansion for the solution we are looking for as follows.

- Let a_0 be the largest integer such that $a_0^2 \leq s$. Why does a_0 exist?
- If $(a_0 + \frac{1}{2})^2 \leq s$, then put $a_1 := a_0 + \frac{1}{2}$, else put $a_1 := a_0$.
- If $(a_1 + \frac{1}{4})^2 \leq s$, then put $a_2 := a_1 + \frac{1}{4}$, else put $a_2 := a_1$.

Continue in this manner to get a Cauchy sequence (a_n) in $\mathbb{Q}$. Let r denote the real number that it represents. Define (b_n) using (2.5). Now by construction,

$$a_n^2 \leq s < b_n^2.$$

Since (a_n) and (b_n) both represent r, (a_n^2) and (b_n^2) both represent r^2. (Explicitly, we can say $b_n^2 - a_n^2 \to 0$ because $b_n^2 - a_n^2 = (b_n - a_n)(b_n + a_n) \leq (b_n - a_n)(2b_0)$.) So by Lemma 2.12, $r^2 \leq s \leq r^2$. Hence $r^2 = s$. (The same proof works in general also.) $\square$

If $r^n = s$, then r is called the positive nth root of s. We denote the positive nth root of s by the symbol $s^{1/n}$. For $s^{1/2}$ we usually write $\sqrt{s}$.

Problems.

(1) Give some examples (other than binary, ternary, ... expansions) of Cauchy sequences in $\mathbb{Q}$ which do not converge in $\mathbb{Q}$. (Look up the wiki page on periodic continued fractions.)
(2) Construct a field which is strictly contains $\mathbb{Q}$ but is strictly contained in $\mathbb{R}$. (Any such field is an example of an ordered field.)
(3) Show that: If $r, s \in \mathbb{R}$ and for each $\epsilon > 0$, $r \leq s + \epsilon$, then $r \leq s$.
(4) Show that: If $r, s \in \mathbb{R}$ and $r > 0$, then there exists a positive integer n such that $nr > s$. This is called the *archimedean property* of real numbers.
(5) A set is uncountable if it is not finite or countable. Show that $\mathbb{R}$ is uncountable.
(6) How does the partial order, multiplication and addition of $\mathbb{R}$ translate to binary expansions?
(7) Show that the rationals correspond to those binary expansions, where a certain finite pattern of 0 and 1 repeats for ever.
(8) Show that between any two distinct irrationals is a rational, and between any two distinct rationals is an irrational.
(9) Consider the following variant of the binary expansion algorithm. Fix a real number r.
    - Let b_0 be the smallest integer such that $b_0 > r$.
    - If $b_0 - \frac{1}{2} > r$, then put $b_1 := b_0 - \frac{1}{2}$ and $e_1 := 1$, else put $b_1 := b_0$ and $e_1 := 0$.
    - If $b_1 - \frac{1}{4} > r$, then put $b_2 := b_1 - \frac{1}{4}$ and $e_2 := 1$, else put $b_2 := b_1$ and $e_2 := 0$.

    Continue in this manner to get a sequence $(b_0, b_1, b_2, \dots)$ in $\mathbb{Q}$ and a sequence $(e_1, e_2, \dots)$ in $\{0, 1\}$. Show that (b_n) coincides with (2.5). What is the relation between the e_n and the digits d_n)?
(10) Show that the least upper bound property or the greatest lower bound does not hold for the poset $\mathbb{Q}$.
(11) Show that: If $r, s > 0$ and n is a positive integer, then $(rs)^{1/n} = r^{1/n}s^{1/n}$. Take a look at [21, Exercises 6 and 7, Chapter 1] to see how exponentials and logarithms can be constructed using the existence of nth roots.
(12) This exercise assumes familiarity with the notion of dimension of a vector space. Show that $\mathbb{R}$ is an infinite-dimensional vector space over $\mathbb{Q}$.
(13) Define the set of complex numbers $\mathbb{C}$ to be $\mathbb{R} \times \mathbb{R}$, that is, a complex number is a pair (x, y) of real numbers. Introduce operations of addition and multiplication on $\mathbb{C}$, and show that $\mathbb{C}$ is a field. Show that the inclusion

$$\mathbb{R} \hookrightarrow \mathbb{C}, \qquad x \mapsto (x, 0)$$

is a ring homomorphism.

2.3. Sequences of real numbers

We discuss sequences in $\mathbb{R}$. The notions of convergence and Cauchyness make sense for such sequences. We show that a sequence in $\mathbb{R}$ converges iff it is Cauchy. We discuss the limit superior and limit inferior of a sequence, and very briefly touch upon series.

2.3.1. Sequences of real numbers. We have constructed the ordered field of real numbers. The material in Section 2.1 is valid with $\mathbb{R}$ in place of $\mathbb{Q}$. Note that ϵ refers to a real number. In particular, we can talk of convergent or Cauchy sequences in $\mathbb{R}$. The absolute value that comes in the defnition is as in (2.3). The properties (2.4) will be routinely used, often implicitly.

Lemma 2.19. *Let r be a real number and (a_n) be any sequence in $\mathbb{Q}$. Then (a_n) is Cauchy and represents r iff (a_n) converges to r.*

PROOF. This is more or less a restatement of Lemma 2.11. □

Lemma 2.20. *Any increasing sequence in $\mathbb{R}$ which is bounded above converges. Dually: any decreasing sequence in $\mathbb{R}$ which is bounded below converges.*

PROOF. We prove the first statement. Let (a_n) be such a sequence. Let a be the least upper bound of the set $\{x \in \mathbb{R} \mid x = a_n \text{ for some } n\}$. (This is the set of distinct values taken by the sequence.) This exists by Proposition 2.16. We claim that $a_n \to a$. Given any $\epsilon > 0$, there is some element say a_N of the sequence greater than $a - \epsilon$. Since the sequence is increasing, this is true for all a_n with $n > N$. The claim follows. □

Proposition 2.21. *Every Cauchy sequence (a_n) in $\mathbb{R}$ converges.*

This is called the Cauchy-completeness property of $\mathbb{R}$. Recall that this property fails in $\mathbb{Q}$, and in a way we have fixed this by enlarging $\mathbb{Q}$ to $\mathbb{R}$.

PROOF. Consider the set

$$\mathrm{A} = \{s \in \mathbb{R} \mid \text{there exist infinitely many } n \text{ such that } s \leq a_n\}.$$

Since any Cauchy sequence is bounded, A is nonempty and bounded above: An upper bound on the sequence gives an upper bound for A, while a lower bound on the sequence yields an element of A. Let r be the least upper bound of A. (Note: r may not be in A. For instance, consider the sequence (a_n) with $a_n := 1 - 1/n$.) We claim that (a_n) converges to r. Let $\epsilon > 0$ be given. We want to show that there is a N such that $|r - a_n| < \epsilon$ for all $n > N$.

Since $r + \epsilon \notin \mathrm{A}$, only finitely many elements of the sequence are greater than it. That is, there exists a N_0 such that $a_n < r + \epsilon$ for all $n > N_0$. We now need to get the other inequality. Use Cauchyness to pick a $N_1 \geq N_0$ such that $|a_m - a_n| < \epsilon/2$ for all $m, n > N_1$. Since $r - \epsilon/2 \in \mathrm{A}$, there is an element a_N with $N > N_1$ which is greater than $r - \epsilon/2$. Combining the two, we deduce that $r - \epsilon < a_n$ for all $n > N$. □

Limits of some special sequences such as $(n^{1/n})$ are discussed in [21, Theorem 3.20] and are for self-study.

2.3.2. Limit superior and limit inferior. Suppose (a_n) is a sequence in the extended real number system $\hat{\mathbb{R}}$. Recall that the latter is a complete lattice, so arbitrary meets and joins exists. It is more common to call them inf and sup. For $k = 0, 1, 2, \ldots$, put

$$b_k := \sup\{a_k, a_{k+1}, a_{k+2}, \ldots\}.$$

Note that $b_0 \geq b_1 \geq b_2 \ldots$. We point out that b_0 and any of the rest can be ∞ or $-\infty$. Put $\beta := \inf\{b_0, b_1, b_2, \ldots\}$. We call β the upper limit or limit superior of (a_n) and write

$$\beta = \limsup_{n\to\infty} a_n.$$

We again emphasize that β could be ∞ or $-\infty$.

The lower limit or limit inferior is defined analogously by interchanging the roles of sup and inf. It is denoted by

$$\liminf_{n\to\infty} a_n.$$

Example 2.22. If a sequence converges, then the upper and lower limits are equal, and equal the limit of the sequence (as we will see below). In the sequence (a_n) with $a_n = (-1)^n$,

$$\limsup_{n\to\infty} a_n = 1 \quad \text{and} \quad \liminf_{n\to\infty} a_n = -1.$$

For the sequence (a_n) with $a_n = n$, the upper and lower limits are both ∞.

By construction,

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

Further, by self-duality of $\mathbb{R}$, we deduce that

$$\liminf_{n\to\infty} a_n = -(\limsup_{n\to\infty} -a_n).$$

2.3.3. Series. Let $\mathsf{Set}(\mathbb{N}, \mathbb{R})$ denote the set of all sequences in $\mathbb{R}$. Consider the map

$$\mathsf{Set}(\mathbb{N}, \mathbb{R}) \to \mathsf{Set}(\mathbb{N}, \mathbb{R}), \qquad (a_n) \mapsto (s_n),$$

where $s_n := a_0 + a_1 + \cdots + a_n$. The sequence (s_n) is called the sequence of partial sums. In fact, given the sequence (s_n), one can recover (a_n) by $a_0 := s_0$ and $a_n := s_n - s_{n-1}$ for $n \geq 1$. Thus the above map is a bijection. It is customary to denote the image of (a_n) under this map by $\sum_{n=0}^{\infty} a_n$, or simply by $\sum a_n$, and call it a *series*. We say that a series $\sum a_n$ converges if its sequence of partial sums (s_n) converges.

A number of ideas related to convergence are convenient to phrase in the language of series. The concepts of absolute convergence, conditional convergence, root test, ratio test, power series, radius of convergence are discussed in detail in [21, pages 59-78] and are for self-study. You may also look at [19, Section 3.3].

Problems.

(1) Show that: If $r_n \to r$ and $s_n \to s$, and $r_n \leq s_n$ for all n, then $r \leq s$. (Here r_n, r, s_n, s are all real numbers.)

(2) Can you prove Proposition 2.21 by working with the set

$$\mathrm{B} := \{s \in \mathbb{R} \mid \text{there exist finitely many } n \text{ such that } a_n < s\}$$

instead of A?

(3) Show that: A sequence (a_n) in $\mathbb{R}$ converges to a iff

$$\liminf a_n = \limsup a_n = a.$$

(4) Show that: If (a_n) is a Cauchy sequence in $\mathbb{R}$, then its upper limit and lower limit coincide and are equal to a real number. Use the previous exercise to deduce Proposition 2.21.

(5) Show that: For any sequence (a_n) in $\mathbb{R}$, there exists a subsequence which converges to $\limsup a_n$. Further, $\limsup a_n$ is the largest number with this property.
(6) Show that: Every bounded sequence in $\mathbb{R}$ has a convergent subsequence. This is known as the Bolzano-Weierstrass theorem.
(7) Let a be an element of the extended real number system $\hat{\mathbb{R}}$. We say that a sequence (a_n) in $\hat{\mathbb{R}}$ converges to a if

$$\liminf a_n = \limsup a_n = a.$$

(This is a definition.) Show that: A sequence (a_n) in $\hat{\mathbb{R}}$ converges to ∞ iff for any real number α, there exists a N such that $a_n > \alpha$ for all $n > N$.
(8) Show that the notions of $\liminf$ and $\limsup$ make sense in any complete lattice. (Thus we can make sense of convergence of a sequence in any complete lattice!) Write explicitly the formulas for the Boolean lattice.
(9) Let X be any set, and A be the set of functions from X to $\mathbb{R}$. For $f, g \in A$, define $f \leq g$ if $f(x) \leq g(x)$ for all $x \in X$. Check that this is a partial order on A. Show that meets and joins exist and describe them. What about arbitrary meets and joins? Can one define $\liminf$ and $\limsup$?

2.4. Continuous functions on the real line

For any extended real numbers $a \leq b$, define

$$(a, b) := \{x \mid a < x < b\} \quad \text{and} \quad [a, b] := \{x \mid a \leq x \leq b\}.$$

Subsets of $\mathbb{R}$ of this kind are called *open intervals* and *closed intervals*, respectively. The notations $(a, b]$ and $[a, b)$ are also used with the obvious meanings. Note that $\mathbb{R} = (-\infty, \infty)$ and $\hat{\mathbb{R}} = [-\infty, \infty]$.

A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous at a point* x_0 if for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$|x - x_0| < \delta \text{ implies } |f(x) - f(x_0)| < \epsilon.$$

A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* if it is continuous at every point x_0.

Continuous functions on open or closed intervals are defined in a similar manner.

A key implication of continuity:

Lemma 2.23. *If f is continuous at x_0, and $f(x_0) < \alpha$, then there is a $\delta > 0$ such that $f(x) < \alpha$ for every x in the interval $(x_0 - \delta, x_0 + \delta)$.*

There is a similar statement for $f(x_0) > \alpha$.

Proof. Take ϵ to be half of $\alpha - f(x_0)$, and apply the definition of continuity. □

In Lemma 2.23, if x_0 is the right endpoint of the domain of f, then the statement needs to be modified slightly by saying: in the interval $(x_0 - \delta, x_0]$. Similar comment for left endpoint.

Lemma 2.24. *A continuous function f on a closed interval $[a, b]$ is bounded. That is, there exist real numbers m and M such that $m \leq f(x) \leq M$ for all $x \in [a, b]$.*

Proof. Consider the set

$$\mathrm{A} = \{x \in [a, b] \mid f([a, x]) \text{ is a bounded subset of } \mathbb{R}\}.$$

This set is nonempty since it contains a, and is bounded above by b. Let c denote its least upper bound, or supremum. By continuity at c, there is an interval around c in which f is bounded. We deduce from this that $c \in \mathrm{A}$ and $c = b$. Can you see this? $\square$

Theorem 2.25. *Suppose f is a continuous function on a closed interval $[a, b]$. Then there exist x_0 and x_1 such that $f(x_0) \leq f(x) \leq f(x_1)$ for all $x \in [a, b]$.*

In other words, a continuous function on a closed interval attains its infimum and supremum values. The general context for this result is compactness as we will see later.

Proof. Let M be the supremum of f on $[a, b]$. (This is the least upper bound of the set $f([a, b])$). It exists by Lemma 2.24. We prove that there exists a x_1 such that $f(x_1) = M$. Consider the set

$$\mathrm{A} = \{x \in [a, b] \mid \text{supremum of } f \text{ on } [a, x] \text{ is strictly smaller than } M\}.$$

If $f(a) = M$, then $x_1 = a$ and we are done. Suppose not. Then A is nonempty since it contains x and is bounded above by b. Let c be its least upper bound. Suppose $f(c) < M$. Then by continuity of f at c, there exists an interval around c where the supremum of the f values is strictly smaller than M. So $c \in \mathrm{A}$. If $c < b$, then there will be numbers greater than c that lie in A which is a contradiction. If $c = b$, then the supremum of f on $[a, b]$ is strictly smaller than M which is again a contradiction. Therefore, $f(c) = M$. So we take $x_1 = c$. $\square$

Theorem 2.26. *Suppose f is a continuous function on $\mathbb{R}$, and a and b are real numbers. Then for any α lying between $f(a)$ and $f(b)$, there exists a c lying between a and b such that $f(c) = \alpha$.*

This is called the intermediate value theorem. The general context for this result is connectedness as we will see later.

Proof. We may assume $a < b$, and further that $f(a) < \alpha < f(b)$. (If $f(b) < f(a)$, then we can replace f by $-f$ to reduce to this case.) Consider the set

$$\mathrm{A} = \{x \in [a, b] \mid \text{supremum of } f \text{ on } [a, x] \text{ is smaller than } \alpha\}.$$

This set is nonempty since it contains a, and is bounded above by b. Let c denote its least upper bound, or supremum. We claim that $f(c) = \alpha$.

- Suppose $f(c) > \alpha$, then there is an open interval around c in which f is always strictly greater than α. In particular, there is a number strictly smaller than c, whose f value is strictly greater than α. This is a contradiction.
- Suppose $f(c) < \alpha$, then there is an open interval around c in which f is always strictly smaller than α. In particular, there is a number strictly greater than c, whose f value is strictly smaller than α. This is a contradiction.

This proves the claim. $\square$

Problems.

(1) Show that $f(x) = x^n$ is a continuous function on $\mathbb{R}$. Deduce Proposition 2.18 using the intermediate value theorem.

(2) Consider the function $f : (0, 1) \to \mathbb{R}$ defined by

$$f(r) := \begin{cases} 0 & \text{if } r \text{ is irrational,} \\ 1/q & \text{if } r \text{ is rational and in lowest terms } r = p/q. \end{cases}$$

Show that f is continuous at every irrational point, but not continuous at any rational point.

(3) Prove the intermediate value theorem by the bisection method. Keep probing at the midpoint of smaller and smaller intervals to converge to the solution.

(4) Show that: If $f : [0, 1] \to [0, 1]$ is continuous, then there exists a x such that $f(x) = x$. (This is a special case of Brouwer's fixed point theorem.)

New section:

Also discuss increasing functions and adjoints ... if time permits.

CHAPTER 3

Topology

We discuss metric spaces, topological spaces, and normed linear spaces. They all provide a context for continuity. Mathematicians use the word "space" to convey the idea of a set with some "geometric" structure. Elements of such a set are usually called points.

3.1. Metric spaces

Roughly speaking, metric spaces are spaces in which there is a notion of distance. The notion of continuity intuitively means that nearby points map to nearby points. This can be formalized in the context of metric spaces. There are many variants of this notion such as uniform continuity, Lipschitz continuity, weak contraction.

Which of these is the correct way to compare metric spaces? This is a philosophical question. Each of one them will give rise to a category. One can perhaps say that topological spaces provide the most general context for continuity, uniform spaces for uniform continuity, and metric spaces for weak contractions, but nothing more.

3.1.1. Metric spaces.

A *metric space* is a set X equipped with a map

$$d : X \times X \to \mathbb{R}$$

such that for all $x, y, z \in X$,

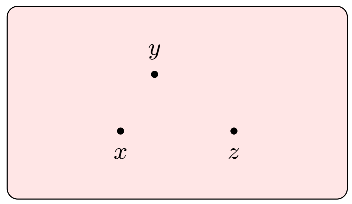
$$\begin{gathered} d(x,x) = 0 \quad \text{and} \quad d(x,y) > 0 \text{ if } x \neq y, \\ d(x,y) = d(y,x), \\ d(x,z) \leq d(x,y) + d(y,z). \end{gathered}$$

The function d is called a *distance function* or a *metric* on X, and $d(x,y)$ is called the distance from x to y. Elements of X are usually called points.

In words: The first property says that the distance from a point to itself is 0, and to any other point is strictly greater than 0. The second property says the distance function is symmetric. The last property is referred to as the triangle inequality motivated from the fact that the sum of two sides of a triangle is greater than the third.

Thus, a metric space is a pair (X, d), where X is a set and d is a metric on X. It is often convenient to keep d implicit and denote the metric space simply by X.

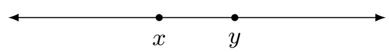
A metric space can be pictured as follows.



X

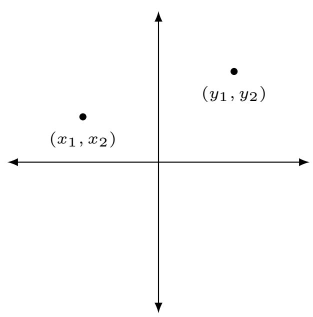
Note: The definition of a metric space involves $\mathbb{R}$ and its structure as an ordered field. This is the reason why we had to first discuss $\mathbb{R}$ before introducing metric spaces.

Example 3.1 (Euclidean metrics). The set of real numbers $\mathbb{R}$ with $d(x, y) := |x - y|$ is a metric space, see (2.4).



Any subset of a metric space is a metric space with the induced metric. For instance, the set of rational numbers $\mathbb{Q}$ with $d(x, y) := |x - y|$ is a metric space, the metric being induced from $\mathbb{R}$.

Consider $\mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$, the cartesian product of $\mathbb{R}$ with itself.



How do we define the distance between points (x_1, x_2) and (y_1, y_2)? There are many interesting choices for a metric on $\mathbb{R}^2$: The euclidean metric

$$d_2((x_1, x_2), (y_1, y_2)) := \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2},$$

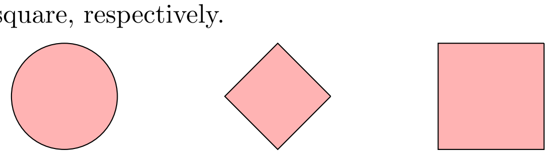
the diamond metric

$$d_1((x_1, x_2), (y_1, y_2)) := |x_1 - y_1| + |x_2 - y_2|,$$

the square metric

$$d_\infty((x_1, x_2), (y_1, y_2)) := \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

are three of the most commonly considered metrics. Check that these are indeed metrics. For these metrics, the set of points at distance 1 from the origin form a circle, diamond, square, respectively.

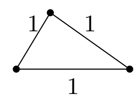


These definitions generalize in an obvious manner to $\mathbb{R}^n$, the n-fold cartesian product of $\mathbb{R}$. There is an even more general context for these metrics where they are called l^2, l^1, l^∞, respectively.

Example 3.2 (Discrete metric). Any set X can be equipped with the metric

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

An illustration is shown below.

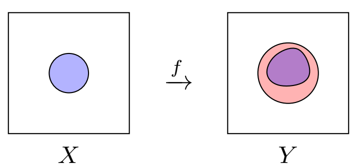


This is called the *discrete metric* on X.

3.1.2. Continuity. Suppose X and Y are metric spaces. A function $f : X \to Y$ is *continuous* if for any point $x_0 \in X$, given $\epsilon > 0$, there exists $\delta > 0$ such that

$$d(x, x_0) < \delta \;\text{ implies }\; d(f(x), f(x_0)) < \epsilon.$$

The d on the left refers to the metric on X, while the one on the right refers to the metric on Y. The condition says that points within distance δ of x_0 map to within distance ϵ of $f(x_0)$.



For $X = Y = \mathbb{R}$ with the usual metric, this definition agrees with our definition of continuous map from $\mathbb{R}$ to $\mathbb{R}$.

Let Metric denote the category whose objects are metric spaces, and whose morphisms are continuous maps. The composition rule is given by usual composition of functions. So we need to check that the composite of two continuous maps is continuous:

Lemma 3.3. *For metric spaces X, Y, Z, if $f : X \to Y$ and $g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is continuous.*

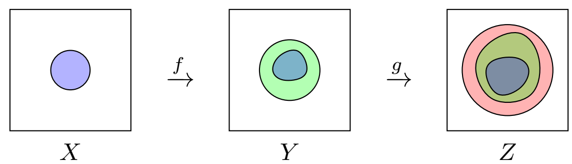
Proof. We show that $g \circ f$ is continuous at each $x_0 \in X$. Let $\epsilon > 0$ be given. By continuity of g at $f(x_0)$, there exists a $\gamma > 0$ such that

$$d(y, f(x_0)) < \gamma \;\text{ implies }\; d(g(y), g(f(x_0))) < \epsilon.$$

Now by continuity of f at x_0, there exists a $\delta > 0$ such that

$$d(x, x_0) < \delta \;\text{ implies }\; d(f(x), f(x_0)) < \gamma.$$

Combining the two yields the required implication. The argument is illustrated below.



∎

The identity map $X \to X$ is clearly continuous, and it serves as the identity morphism id_X in the category. Since metric spaces are sets with structure, there is a forgetful functor $\mathsf{Metric} \to \mathsf{Set}$.

3.1.3. Uniform continuity. A function $f : X \to Y$ between metric spaces is *uniformly continuous* if given $\epsilon > 0$, there exists $\delta > 0$ such that

$$d(x, y) < \delta \ \text{ implies } \ d(f(x), f(y)) < \epsilon.$$

It is clear that a uniformly continuous function is continuous, but the converse is not true in general. The term uniform is used because given ϵ, a δ can be chosen independent of x_0.

Lemma 3.4. *The composite of two uniformly continuous maps is uniformly continuous.*

Proof. Suppose $f : X \to Y$ and $g : Y \to Z$ are uniformly continuous. We show that $g \circ f$ is uniformly continuous. Let $\epsilon > 0$ be given. By uniform continuity of g, there exists a $\gamma > 0$ such that

$$d(f(x), f(y)) < \gamma \ \text{ implies } \ d(g(f(x)), g(f(y))) < \epsilon.$$

Now by uniform continuity of f, there exists a $\delta > 0$ such that

$$d(x, y) < \delta \ \text{ implies } \ d(f(x), f(y)) < \gamma.$$

Combining the two yields the required implication. ∎

Let Metric_u denote the category whose objects are metric spaces, and morphisms are uniformly continuous maps. Since any uniformly continuous map is also continuous, Metric_u is a subcategory of Metric (in which we have kept all objects, but thrown away some morphisms). Note: Being uniformly continuous is a property of a continuous map.

3.1.4. Lipschitz continuity. A function $f : X \to Y$ between metric spaces is *Lipschitz continuous* if there exists a $K > 0$ such that

$$d(f(x), f(y)) \leq K\, d(x, y)$$

for all $x, y \in X$.

If the above holds, then K is called a Lipschitz constant for f.

Observe that the composite of two Lipschitz continuous maps is Lipschitz continuous: With notations as before, for all $x, y \in X$,

$$d(g(f(x)), g(f(y))) \leq K'\, d(f(x), f(y)) \leq KK'\, d(x, y),$$

where K is a Lipschitz constant for f, and K' is a Lipschitz constant for g.

Thus, we obtain the category Metric_L whose objects are metric spaces and morphisms are Lipschitz continuous maps.

3.1.5. Weak contraction. A function $f : X \to Y$ between metric spaces is a *weak contraction* or a *short map* or a *nonexpansive map* if

$$d(f(x), f(y)) \leq d(x, y)$$

for all $x, y \in X$. It is clear that the composite of two weak contractions is a weak contraction. Thus, we obtain the category Metric_{wc} whose objects are metric spaces and morphisms are weak contractions.

3.1.6. Bounded metric spaces. A metric space (X, d) is *bounded* if there exists a real number M such that $d(x, y) < M$ for all $x, y \in X$. More generally, a subset A of X is bounded if it is bounded under the induced metric. For a nonempty bounded set A,

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

is called the *diameter* of A.

Example 3.5. A single point in any metric space is bounded with diameter 0. Any discrete metric space is bounded, and any subset with more than one point has diameter 1. The real line $\mathbb{R}$ in the usual metric is not bounded, but the subsets $(0, 1)$ and $[0, 1]$ are bounded with diameter 1.

Boundedness is a property of a metric space. So we can consider the full subcategory of Metric whose objects are bounded metric spaces. Let us denote it by $\mathsf{BoundedMetric}$. Interestingly, there is no loss of information if we restrict to this subcategory. Formally, one says that the categories $\mathsf{BoundedMetric}$ and Metric are equivalent via the inclusion functor. The content of this is explained below.

Suppose d is a metric on X. Define

$$\bar{d}(x, y) = \min\{d(x, y), 1\}.$$

Then $\bar{d}$ is also a metric on X, and by construction, it is bounded by 1. It is called the *standard bounded metric* associated to d. In fact, the identity maps $(X, d) \to (X, \bar{d})$ and $(X, \bar{d}) \to (X, d)$ are both continuous. Check this. This shows that (X, d) and $(X, \bar{d})$ are isomorphic objects in Metric. Thus every metric space is isomorphic to a bounded metric space.

There is a stronger property than boundedness called total boundedness. At the moment, we only give its definition.

A metric space X is *totally bounded* if for every $\epsilon > 0$, X has a finite open cover by balls of radius ϵ. We will refer to these as ϵ-balls.

Problems.

(1) Show that: For any points a, b, x, y in a metric space X,

$$|d(a, b) - d(x, y)| \leq d(a, x) + d(b, y). \tag{3.1}$$

(2) Show that: The three metric spaces, namely, $\mathbb{R}^2$ with the euclidean, diamond, square metrics, are all isomorphic in Metric (under the identity map!). (This allows us to talk of a continuous map from $\mathbb{R}^2$ to $\mathbb{R}$ without ambiguity.)

(3) Show that addition and multiplication on $\mathbb{R}$ are continuous functions from $\mathbb{R}^2 \to \mathbb{R}$. Deduce that the sum and product of real-valued continuous functions on a metric space X are continuous.

(4) Check that the diamond and square metrics on $\mathbb{R}^n$ are indeed metrics. Show that the euclidean metric on $\mathbb{R}^n$ is indeed a metric. (The triangle inequality in this context is equivalent to Minkowski's inequality.)
(5) For any metric spaces X and Y, put three metrics on $X \times Y$ by analogy with the euclidean, diamond, square metrics on $\mathbb{R}^2$. Show that: For any metric space X, the distance function $d : X \times X \to \mathbb{R}$ is continuous (wrt either of the above three metrics on $X \times X$).
(6) Describe the initial object, terminal object, product, coproduct in the category Metric.
(7) Check that the assignment which sends a set X to the set X with the discrete metric gives rise to a functor

$$\mathsf{Set} \to \mathsf{Metric}.$$

Show that it is the left adjoint to the forgetful functor from Metric to Set. What about a right adjoint?
(8) Check that: A weak contraction is Lipschitz continuous, a Lipschitz continuous map is uniformly continuous, a uniformly continuous map is continuous. In other words, we have functors

$$\mathsf{Metric}_{wc} \to \mathsf{Metric}_L \to \mathsf{Metric}_u \to \mathsf{Metric}.$$

Show that all inclusions are proper, that is, there are continuous maps which are not uniformly continuous, and so on.
(9) A function $X \to Y$ between metric spaces is an *isometry* if it preserves distances, that is, $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$. For instance, the map

$$\mathbb{R} \to \mathbb{R}^2, \qquad t \mapsto \big(\frac{3}{5}t + 1, \frac{4}{5}t - 5\big)$$

is an isometry, with the image being the line $4x = 3y + 19$. In which of the metric categories is bijective isometry the notion of isomorphism?
(10) A function $X \to Y$ between metric spaces is *bi-Lipschitz continuous* if there exists a $K > 0$ such that

$$\frac{1}{K}\, d(x, y) \leq d(f(x), f(y)) \leq K\, d(x, y).$$

for all $x, y \in X$. In which of the metric categories is bijective bi-Lipschitz continuous the notion of isomorphism?
(11) What is wrong with the following argument which shows that any continuous function is uniformly continuous?
Suppose f is a continuous function from X to Y. Let $\epsilon > 0$ be given. For each $x_0 \in X$, pick a $\delta_{x_0} > 0$ given by continuity at x_0. Define δ to be the minimum of all the δ_{x_0}, and this δ works.
(12) Show that: If a metric space is totally bounded, then it is bounded. The converse is not true in general.

3.2. Convergent and Cauchy sequences

We consider sequences in a metric space. By virtue of the metric, the notion of convergence and Cauchyness make sense. We discuss these notions, compare with the discussion in Section 2.1. (There is no addition or multiplication in an arbitrary metric space, so we cannot add or multiply sequences.)

3.2.1. Sequences. A sequence (a_n) in X, that is, $a_n \in X$ for all n, is *convergent* if there exists a $a \in X$ such that for every real number $\epsilon > 0$, there exists an integer N such that $d(a_n, a) < \epsilon$ for all $n > N$. The point a is called the *limit* of the sequence (a_n), and we say that (a_n) converges to a. We denote it by $a_n \to a$ or

$$\lim_{n\to\infty} a_n = a.$$

Lemma 3.6. *The limit of a sequence, if it exists, is unique.*

Proof. Suppose $a_n \to a$ and $a_n \to b$ and $a \neq b$. Choose a real number $\epsilon > 0$ smaller than half of $d(a, b)$. There exists an integer N such that for all $n > N$, $d(a_n, a) < \epsilon$ and $d(a_n, b) < \epsilon$. For any such n, employing symmetry and the triangle inequality, we obtain

$$d(a, b) \leq d(a_n, a) + d(a_n, b) \leq 2\epsilon < d(a, b),$$

which is a contradiction. □

The notation $a_n \not\to a$ means that (a_n) does not converge to a. (In this case, the sequence is either not convergent or converges to some other point.) The statement $a_n \not\to a$ is equivalent to: There exists a real number $\epsilon > 0$ and a subsequence (a_{i_n}) such that $d(a_{i_n}, a) > \epsilon$ for all n.

A sequence (a_n) in X is *Cauchy* if for every real number $\epsilon > 0$, there exists an integer N such that $d(a_m, a_n) < \epsilon$ for all $m, n > N$.

Lemma 3.7. *If a sequence in X is convergent, then it is Cauchy.*

Proof. Suppose (a_n) is a sequence which converges to a. Let $\epsilon > 0$ be given. Choose N such that $d(a_n, a) < \epsilon/2$ for all $n > N$. Now for any $m, n > N$,

$$d(a_m, a_n) \leq d(a_m, a) + d(a_n, a) < \epsilon/2 + \epsilon/2 = \epsilon.$$

So (a_n) is Cauchy. □

3.2.2. Continuity via sequences. There is a nice characterization of continouus functions in terms of convergent sequences:

Proposition 3.8. *Let $f : X \to Y$ be a function between metric spaces. Then f is continuous iff for every $a_n \to a$ in X, we have $f(a_n) \to f(a)$ in Y.*

Proof. Forward implication. Suppose $a_n \to a$. Let $\epsilon > 0$ be given. Choose $\delta > 0$ such that

$$d(x, a) < \delta \text{ implies } d(f(x), f(a)) < \epsilon.$$

Now use this δ to pick N such that for all $n > N$, $d(a_n, a) < \delta$. Then $d(f(a_n), f(a)) < \epsilon$ for all $n > N$, and hence $f(a_n) \to f(a)$.

Backward implication. Suppose f fails to be continuous at a. Then there is a ϵ for which no δ works. Thus, for each n, there is a a_n such that $d(a_n, a) < 1/n$ but $d(f(a_n), f(a)) > \epsilon$. Then by construction $a_n \to a$, but $f(a_n) \not\to f(a)$. □

Problems.

(1) Show that: If X and Y are isomorphic in Metric, then there is a bijection between the set of convergent sequences in X and the set of convergent sequences in Y.

(2) Let X and Y be metric space, and $X \times Y$ be the metric space with the square metric. Show that: A sequence (a_n, b_n) in $X \times Y$ is Cauchy iff (a_n) is Cauchy in X and (b_n) is Cauchy in Y. Similarly, $(a_n, b_n) \to (a, b)$ in $X \times Y$ iff $a_n \to a$ in X and $b_n \to b$ in Y. Is this true for the euclidean and diamond metrics?
(3) A sequence is a metric space X is *bounded* if its image is a bounded subset of X. Show that any Cauchy sequence (and hence any convergent sequence) in a metric space is bounded.
(4) Show that: Given any point x and a bounded subset A in a metric space X, there exists an r such that A is a subset of $B(x, r)$.

3.3. Complete metric spaces

A metric space is complete if every Cauchy sequence in it converges. Every metric space can be completed in a canonical manner. This gives rise to the completion functor from the category of metric spaces to the category of complete metric spaces.

3.3.1. Complete metric spaces. A metric space X is called *complete* if every Cauchy sequence in X converges. This is a property of a metric space. For example, $\mathbb{R}$ is complete while $\mathbb{Q}$ is not (under usual metrics). Thus, we have the full subcategory $\mathsf{CompleteMetric}$ of Metric whose objects are complete metric spaces, and morphisms are continuous maps. This subcategory is full because whenever two objects are kept all morphisms between them are kept as well.

Similarly, we have the subcategory $\mathsf{CompleteMetric}_u$ of Metric_u whose objects are complete metric spaces, and morphisms are uniformly continuous maps, and so on.

3.3.2. Completion functor. There is a canonical way to complete a metric space by copying the construction of $\mathbb{R}$ starting with $\mathbb{Q}$. Details are as follows.

Fix a metric space (X, d). Let Ω denote the set of all Cauchy sequences in X. Define an equivalence relation on Ω by:

$$(a_n) \sim (b_n) \quad \text{if} \quad \lim_{n \to \infty} d(a_n, b_n) = 0. \tag{3.2}$$

The condition means that given any real number $\epsilon > 0$, there exists an integer N such that $d(a_n, b_n) < \epsilon$ for all $n > N$.

We check that this is an equivalence relation.

- $d(a_n, a_n) = 0$, hence reflexive.
- $d(a_n, b_n) = d(b_n, a_n)$, hence symmetric.
- $0 \leq d(a_n, c_n) \leq d(a_n, b_n) + d(b_n, c_n)$, hence transitive.

(Note how the three axioms of a metric, respectively, imply the three conditions of an equivalence relations.)

Observation 3.9. If (a_n) converges to a, then every subsequence of (a_n) also converges to a. Similarly, if (a_n) is a Cauchy sequence, then every subsequence of (a_n) is also a Cauchy sequence and $\sim$-equivalent to (a_n).

Define $\hat{X}$ to be the set of equivalence classes $\Omega_\sim$. Thus, an element of $\hat{X}$ is an equivalence class of Cauchy sequences in X.

Every $a \in X$ gives rise to the constant Cauchy sequence $(a, a, a, \dots)$. This yields a map

$$X \hookrightarrow \hat{X}, \qquad a \mapsto [(a, a, a, \dots)]. \tag{3.3}$$

The square brackets around the sequence are used to indicate the equivalence class of the sequence. This map is injective: Suppose $(a, a, \dots) \sim (b, b, \dots)$. Then $d(a, b) \to 0$. Since this is a constant sequence, $d(a, b) = 0$. Hence $a = b$.

We proceed to define a metric

$$D : \hat{X} \times \hat{X} \to \mathbb{R}.$$

Given points $a, b \in \hat{X}$, pick representatives (a_n) and (b_n). Using the estimate (3.1)

$$|d(a_m, b_m) - d(a_n, b_n)| \leq d(a_m, a_n) + d(b_m, b_n),$$

we see that $(d(a_n, b_n))$ is a Cauchy sequence in $\mathbb{R}$. By completeness of $\mathbb{R}$ (Proposition 2.21), it converges to some real number, say r. Put $D(a, b) = r$.

We need to check that D is well-defined. Accordingly, suppose $(a_n) \sim (a'_n)$ and $(b_n) \sim (b'_n)$. Then

$$|d(a_n, b_n) - d(a'_n, b'_n)| \leq d(a_n, a'_n) + d(b_n, b'_n).$$

Since both terms in the rhs go to 0, so does the lhs. Hence $d(a_n, b_n)$ and $d(a'_n, b'_n)$ converge to the same real number, and D is well-defined.

Lemma 3.10. *D is a metric.*

Proof. We need to check the three conditions of a metric.

- $D(a, b) \geq 0$, since $D(a, b)$ is the limit of a sequence $(d(a_n, b_n))$ of nonnegative terms. Further, $D(a, b) = 0$ iff $d(a_n, b_n) \to 0$ iff $(a_n) \sim (b_n)$ iff $a = b$.
- For $D(a, b) = D(b, a)$: By symmetry of d, $d(a_n, b_n) = d(b_n, a_n)$. Now let n go to infinity.
- For $D(a, c) \leq D(a, b) + D(b, c)$: By triangle inequality for d, $d(a_n, c_n) \leq d(a_n, b_n) + d(b_n, c_n)$. Now let n go to infinity.

□

Lemma 3.11. *The metric space $(\hat{X}, D)$ is complete.*

Proof. Let (A_k) be a Cauchy sequence in $\hat{X}$. In particular, $A_k \in \hat{X}$. Represent each element A_k by a Cauchy sequence $(a_{k,n})$ in X with the property that

$$d(a_{k,m}, a_{k,n}) < \frac{1}{k}$$

for all m, n. (This can be done by picking any representative $(a'_{k,n})$ and then deleting the first few terms so that the remaining terms differ from one another by less that $1/k$. The resulting subsequence also represents A_k by Observation 3.9.)

Now consider the sequence (b_n) in X defined by $b_n := a_{n,n}$. We claim that (b_n) is a Cauchy sequence in X. For any k, l, n, we have

$$d(b_k, b_l) = d(a_{k,k}, a_{l,l}) \leq d(a_{k,k}, a_{k,n}) + d(a_{l,n}, a_{l,l}) + d(a_{k,n}, a_{l,n}) < \frac{1}{k} + \frac{1}{l} + d(a_{k,n}, a_{l,n}).$$

Letting n go to infinity (what does this mean?), we get

$$d(b_k, b_l) \leq \frac{1}{k} + \frac{1}{l} + D(A_k, A_l).$$

Since (A_k) is Cauchy, we deduce that (b_n) is Cauchy. Can you write down the formal argument? Let B denote the equivalence class of (b_n). We claim that (A_k)

converges to B, that is $D(A_k, B) \to 0$. So we need to estimate the distance d between the nth terms of the sequences that represent A_k and B:

$$d(a_{k,n}, b_n) \leq d(a_{k,n}, b_k) + d(b_k, b_n) < \frac{1}{k} + d(b_k, b_n).$$

Given $\epsilon' > 0$, there is a N_0 such that for all $k, n > N_0$, $d(b_k, b_n) < \epsilon'$. So for any $k > N_0$, by letting n go to infinity, we obtain

$$D(A_k, B) \leq \epsilon' + \frac{1}{k}.$$

Suppose $\epsilon > 0$ is given. There is a N_1 such that for all $k > N_1$, $1/k < \epsilon/2$. Using the displayed inequality for $\epsilon' = \epsilon/2$, we deduce that $D(A_k, B) \leq \epsilon$ for all $k > N$, where $N = \max\{N_0, N_1\}$. □

Suppose $f : X \to Y$ is uniformly continuous. If (a_n) is a Cauchy sequence in X, then $(f(a_n))$ is a Cauchy sequence in Y: Suppose $\epsilon > 0$ is given. Uniform continuity of f yields a $\delta > 0$. Pick N such that $d(a_m, a_n) < \delta$ for all $m, n > N$. Then $d(f(a_m), f(a_n)) < \epsilon$, as required. Further, if $(a_n) \sim (b_n)$ as defined in (3.2), then $(f(a_n)) \sim (f(b_n))$. This induces a map

$$\hat{f} : \hat{X} \to \hat{Y},$$

and in fact $\hat{f}$ is uniformly continuous. Check these details. In other functor, we have a functor

$$\mathsf{Metric}_u \to \mathsf{CompleteMetric}_u, \qquad X \mapsto \hat{X}.$$

We may call this the *completion functor*.

There is another standard construction of the completion, see [18, Section 43].

Problems.

(1) Show that: If X is complete, then (3.3) is a bijection.
(2) Show that a continuous map between metric spaces does not preserve Cauchy sequences in general. That is, the completion construction is not functorial wrt continuous maps.
(3) Show that: If X and Y are complete, then so is $X \times Y$ (say wrt to the square metric). Check that this is the categorical product in $\mathsf{CompleteMetric}$.
(4) Which subsets of $\mathbb{R}$ (with metric induced from $\mathbb{R}$) are complete?
(5) Show that $\mathsf{CompleteMetric}$ is not an isomorphism-closed subcategory of Metric, but $\mathsf{CompleteMetric}_u$ is an isomorphism-closed subcategory of Metric_u. Explicitly, show that there are two homeomorphic metric spaces such that one is complete while the other is not. (Books often say: Completeness is not a topological property. This is the same as the first statement.)
(6) Show that the completion functor is the left adjoint to the inclusion functor. In other words, a uniformly continuous map $f : X \to Y$, where Y is complete, extends uniquely to a uniformly continuous map $\hat{f} : \hat{X} \to Y$.
(7) Give an example of a sequence of Cauchy sequences in a metric space X whose 'diagonal' is not a Cauchy sequence.
(8) Let X be a metric space. Show that: X is complete iff every Cauchy sequence in X has a convergent subsequence.

3.4. Topological spaces

Roughly speaking, topological spaces are spaces in which there is a qualitative notion of closeness. This is formalized through the notion of open sets. This is in contrast to metric spaces where we have a quantitative notion of closeness.

3.4.1. Topological spaces. A *topological space* is a set X equipped with a collection τ of subsets of X (that is, τ is a subset of 2^X) such that the following axioms hold.

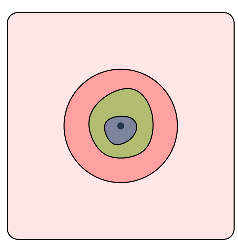
- $\emptyset$ and X are in τ.
- the union of the elements in any subset of τ is in τ.
- the intersection of the elements in any finite subset of τ is in τ.

The collection τ is called a topology on X, and its elements are called *open sets*. The above axioms can be rephrased using the language of open sets as follows.

- The empty set and the full space are open.
- Arbitrary union of open sets is open.
- Finite intersection of open sets is open.

Thus, a topological space is a pair (X, τ) consisting of a set and a topology on it. However, it is often convenient to simply write X and keep the topology τ implicit. This is similar to the practice for metric spaces where the metric d is kept implicit.

A topological space can be pictured similar to a metric space. The difference is that instead of distance between points we now have open sets containing points.

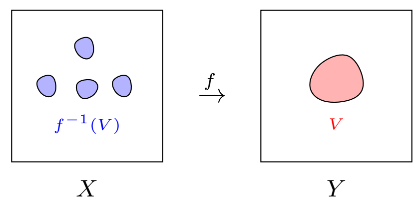


The picture shows smaller and smaller open sets containing a point. These indicate how close we can get to that point.

Example 3.12. Let X be any set. The collection of all subsets of X is a topology on X, that is, $\tau = 2^X$. It is called the *discrete topology*. The collection $\tau = \{\emptyset, X\}$ is also a topology on X. It is called the *indiscrete topology*.

In general, given any subset S of 2^X, there is a smallest topology τ on X which contains S: The discrete topology on X contains S. Now take the (nonempty) intersection of all topologies which contains S. Check that this is a topology on X. By construction, this is the smallest topology on X which contains S. In this situation, τ is called the topology *generated* by S and S is called a *subbase* of τ.

3.4.2. Continuity. Suppose X and Y are topological spaces. A function $f : X \to Y$ is *continuous* if for any open set V of Y, $f^{-1}(V)$ is an open set of X.



Let Top denote the category whose objects are topological spaces, and whose morphisms are continuous maps. The composition rule is given by usual composition of functions. We check:

Lemma 3.13. *For topological spaces X, Y, Z, if $f : X \to Y$ and $g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is continuous.*

Proof. Suppose U is open in Z. Since g is continuous, $g^{-1}(U)$ is open in Y. Now since f is continuous, $f^{-1}(g^{-1}(U))$ is open in X. But $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U)$, and the result follows. □

By construction, we have a forgetful functor Top $\to$ Set. Now, a topological space is a set with more structure, which is why we use the term forgetful functor. However, continuity is a property of a function. Hence this functor is faithful: distinct continuous maps have distinct underlying functions.

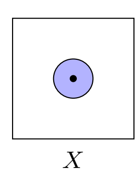
A continuous map $f : X \to Y$ between topological spaces is a *homeomorphism* if it is an isomorphism in the category Top.

Explicitly, $f : X \to Y$ is a homeomorphism if there is a continuous map $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$.

3.4.3. Underlying topology of a metric space. Let X be a metric space. The subset

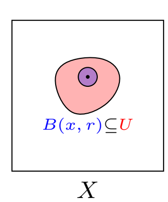
$$B(x, r) := \{y \in X \mid d(x, y) < r\}$$

is called the open ball with center at x and radius r.



Note that if $r \leq s$, then $B(x, r) \subseteq B(x, s)$.

A subset U of X is *open* if for every $x \in U$, there exists a $r > 0$ such that $B(x, r) \subseteq U$.

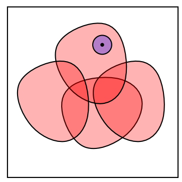


By the triangle inequality, one can see that any open ball is indeed open. This explains the terminology "open ball".

Lemma 3.14. *The collection of open sets of a metric space X defines a topology on X.*

PROOF. We need to verify the axioms for a topology.

- Clearly, $\emptyset$ and X are open.
- Suppose $\{U_i\}_{i \in I}$ is an arbitrary collection of open sets. (The indexing set I may not be finite or even countable.) Then the union of the U_i is also open: Suppose x is in this union. Then it must belong to at least one member of the collection. Say $x \in U_j$. By openness of U_j, there exists a $r > 0$ such that $B(x, r) \subseteq U_j$. Since the union is larger than U_j, the ball $B(x, r)$ is also a subset of the union.



- Similarly, argue for intersections. Why do we have to restrict to a finite indexing set now?

□

Thus, to every metric space (X, d), one can associate the topological space (X, τ_d), where τ_d is the collection of open sets of (X, d). We call τ_d the underlying topology of d.

Lemma 3.15. *Suppose X and Y are metric spaces and $f : X \to Y$ is a function. Then f is continuous iff for every open set V in Y, $f^{-1}(V)$ is an open set of X.*

PROOF. Forward implication. Let V be an open set in Y. We want to show that $f^{-1}(V)$ is an open set of X. Take $x_0 \in f^{-1}(V)$. Hence $f(x_0) \in V$. Since V is open, choose $\epsilon > 0$ such that $B(f(x_0), \epsilon) \subseteq V$. Continuity at x_0 yields a $\delta > 0$ such that

$$B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \epsilon)) \subseteq f^{-1}(V).$$

Thus, $f^{-1}(V)$ is open in X.

Backward implication. We show that f is continuous at each $x_0 \in X$. We are given $\epsilon > 0$. Consider $B(f(x_0), \epsilon)$. This is an open set in Y. By hypothesis, its inverse image is open in X and contains x_0. Therefore, there exists a $\delta > 0$ such that $B(x_0, \delta) \subseteq f^{-1}(B(f(x_0), \epsilon))$. This is precisely the continuity condition. □

Proposition 3.16. *There is a full and faithful functor*

$$\mathsf{Metric} \to \mathsf{Top}. \tag{3.4}$$

PROOF. This is a consequence of Lemmas 3.14 and 3.15: The functor sends a metric space (X, d) to (X, τ_d), and a continuous function f between two metric spaces to f itself. Denoting this functor by $\mathcal{F}$, the map

$$\mathsf{Metric}(X, Y) \to \mathsf{Top}(\mathcal{F}(X), \mathcal{F}(Y)), \qquad f \mapsto f$$

is clearly a bijection. This is why $\mathcal{F}$ is full and faithful. □

A topological space (X, τ) is called *metrizable* if there exists a metric d on X whose underlying topology is τ. Thus Metric is equivalent to the full subcategory of metrizable topological spaces.

Example 3.17 (Standard topology on euclidean space). Let us begin with $\mathbb{R}$. It is a metric space under the usual metric. So it has an underlying topology. A set is open in this topology if it can be written as a union of open intervals. In particular, any open interval is an open set. In fact, any open interval is an instance of an open ball in the metric (with center at the midpoint of the interval). Observe that

$$\bigcap_{n\geq 1} (\frac{-1}{n}, \frac{1}{n}) = \{0\}.$$

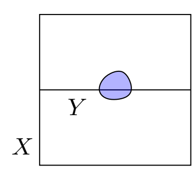
Now $\{0\}$ is not an open set. Hence this identity shows that an arbitrary intersection (even countable intersection) of open sets is not open in general.

In general, $\mathbb{R}^n$ has an underlying topology wrt the euclidean metric. We call it the *standard topology* on $\mathbb{R}^n$. A set is open in this topology if it can be written as a union of open balls. What happens if we use the diamond metric or square metric instead?

3.4.4. Subspace topology. Suppose (X, τ) is a topological space. Let Y be any subset of X. This can be viewed as an injective map

$$i : Y \hookrightarrow X.$$

We now put a topology on Y. Call a subset V of Y open if there exists an open set U in X such that $V = U \cap Y$.



Let τ_Y denote the collection of these open sets. That is,

$$\tau_Y = \{V \subseteq Y \mid V = U \cap Y \text{ for some } U \in \tau\}.$$

We claim that τ_Y is a topology on Y. It is called the *subspace topology*.

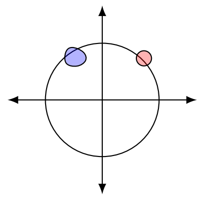
- First, $\emptyset$ and Y are open since
$$\emptyset = \emptyset \cap Y \quad \text{and} \quad Y = X \cap Y,$$
and $\emptyset$ and X are open in X.
- To verify that arbitrary union of open sets is open, suppose that for each $i \in I$, we are given a $V_i \in \tau_Y$. Then $V_i = U_i \cap Y$ for some $U_i \in \tau$. Now
$$\bigcup_{i\in I} V_i = \bigcup_{i\in I} U_i \cap Y = (\bigcup_{i\in I} U_i) \cap Y.$$
Since τ is a topology, $\cup_{i\in I} U_i \in \tau$, and hence $\cup_{i\in I} V_i \in \tau_Y$ as required.
- Similarly, the identity
$$V_1 \cap \cdots \cap V_k = (U_1 \cap Y) \cap \cdots \cap (U_k \cap Y) = (U_1 \cap \cdots \cap U_k) \cap Y$$
shows that finite intersections of open sets is open.

Observe that the inclusion map i is continuous with this topology on Y.

Example 3.18 (Spheres). The unit circle is defined by

$$S^1 := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}.$$

It is a subset of $\mathbb{R}^2$. We view it as a topological space with the subspace topology.



More generally, the unit n-sphere defined by

$$S^n := \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}$$

is a topological space with subspace topology inherited from $\mathbb{R}^{n+1}$.

Example 3.19. Any open interval (a, b) with $a < b$ is a topological space with subspace topology inherited from $\mathbb{R}$. By translation and scaling, we see that (a, b) is homeomorphic to $(-1, 1)$. Moreover,

$$(-1, 1) \to \mathbb{R}, \qquad y \mapsto \frac{y}{1 - |y|}$$

is an homeomorphism. The formula for the inverse is $z \mapsto \frac{z}{1+|z|}$. Thus, any open interval (a, b) with $a < b$ is homeomorphic to $\mathbb{R}$.

Similarly, e^x is a homeomorphism from $\mathbb{R}$ to $(0, \infty)$ with inverse given by $\log(x)$.

More generally: The open n-ball of radius r around x_0 is

$$B(x_0, r) := \{x \in \mathbb{R}^n \mid x - x_0 < r\}.$$

We give it the subspace topology from $\mathbb{R}^n$.

By translation and scaling, we see that $B(x_0, r)$ and $B(0, 1)$ are homeomorphic. Further, the map

$$B(0, 1) \to \mathbb{R}^n, \qquad y \mapsto \frac{y}{1 - y} \tag{3.5}$$

is an homeomorphism. The formula for the inverse is $z \mapsto \frac{z}{1+z}$.

3.4.5. Quotient topology. Suppose (X, τ) is a topological space. Let Y be a set and

$$p : X \twoheadrightarrow Y$$

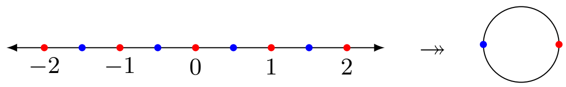
a surjective map. This is the same as specifying an equivalence relation on X whose equivalence classes are elements of Y. We now put a topology on Y. Call a subset V of Y open if $p^{-1}(V)$ is open in X. This is called the *quotient topology.* Observe that the map p is continuous with this topology on Y.

Example 3.20 (Circle). Let $X = [0, 1]$ with subspace topology from $\mathbb{R}$. Now consider the quotient space Y obtained from X by identifying the points 0 and 1. Can we identify Y with some known topological space? Yes, $Y = S^1$ with the quotient map illustrated below.

A related idea is to view the circle as a quotient space of the real line via the map

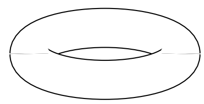
$$p : \mathbb{R} \twoheadrightarrow S^1, \qquad x \mapsto e^{2\pi i x}$$

as illustrated below.

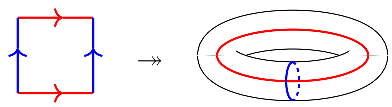


This is an example of what is called a covering map. The group of integers $\mathbb{Z}$ acts on $\mathbb{R}$ by translation $m \cdot x := x + m$ and orbits of this action are precisely points of S^1. Strictly speaking: Group action is not defined in these notes.

Example 3.21 (Torus). Let $X = [0,1] \times [0,1]$ with subspace topology from $\mathbb{R}^2$. Now identify the two horizontal lines on the boundary $(x, 0) \sim (x, 1)$ and the two vertical lines on the boundary $(0, y) \sim (1, y)$. The resulting quotient space Y is called the torus.

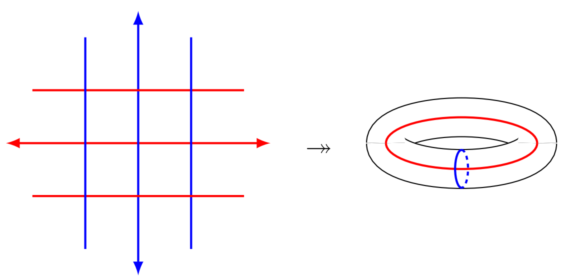


The quotient map is illustrated below.



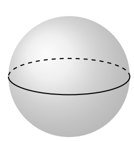
The torus can also be interpreted as $S^1 \times S^1$.

Following the idea from the previous example, the torus can be viewed as a quotient of $\mathbb{R}^2$ in which we identify two points whenever they differ by an integer in either coordinate:



This is another example of a covering map. The group $\mathbb{Z}^2$ acts on $\mathbb{R}^2$ by translation $(m, n) \cdot (x, y) := (x + m, y + n)$ and orbits of this action are points of the torus.

Example 3.22 (Real projective space). Let $\mathbb{R}P^2$ denote the quotient of S^2 obtained by identifying antipodal points, that is, $x \sim -x$. This is called the real projective plane.



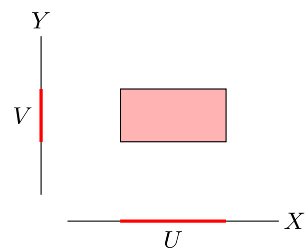
$$\twoheadrightarrow \quad \mathbb{R}P^2$$

More generally: Recall the n-sphere S^n. Let $\mathbb{R}P^n$ denote the quotient of S^n obtained by identifying antipodal points. This is called the n-dimensional real projective space. The quotient map $S^n \twoheadrightarrow \mathbb{R}P^n$ is also a covering map. The group $\mathbb{Z}_2$

with two elements acts on S^n with the nontrivial element acting by $x \mapsto -x$ and orbits of this action are points of $\mathbb{R}P^n$.

3.4.6. Categorical product. Product topology. Let X and Y be topological spaces. We define a topology on the cartesian product $X \times Y$ as follows.

An open rectangle in $X \times Y$ is a subset of the form $U \times V$, where U is an open set in X and V is an open set in Y. An illustration is shown below.



Note that intersection of two open rectangles is again an open rectangle. A subset W of $X \times Y$ is open if it can be written as the union of open rectangles. In other words, for each point $w \in W$, there exists an open rectangle $U \times V$ such that $U \times V \subseteq W$ and $w \in U \times V$. In particular, any open rectangle is an open set.

This defines a topology on $X \times Y$ called the *product topology.* Let us check the axioms.

- $\emptyset$ and $X \times Y$ are open rectangles, hence open.
- Suppose for each $i \in I$, W_i is open. Then each W_i is an union of open rectangles, and hence so is $\cup_{i \in I} W_i$. This shows that the latter is open.
- Suppose W and W' are open. Then write $W = \cup_i U_i \times V_i$ and $W' = \cup_j U'_j \times V'_j$. Hence
$$W \cap W' = \big(\bigcup_i U_i \times V_i\big) \cap \big(\bigcup_j U'_j \times V'_j\big) = \bigcup_{i,j} (U_i \cap U'_j) \times (V_i \cap V'_j)$$
is a union of open rectangles, and open.

The canonical projections
$$\pi_1 : X \times Y \to X \quad \text{and} \quad \pi_2 : X \times Y \to Y$$
are continuous: For any open set U in X, $\pi_1^{-1}(U) = U \times Y$, which is an open rectangle and hence open in $X \times Y$. This shows that π_1 is continuous. The argument for π_2 is similar.

Suppose Z is another topological space, and $f : Z \to X$ and $g : Z \to Y$ are continuous maps. Then we claim that
$$h : Z \to X \times Y, \qquad h(z) = (f(z), g(z))$$
is continuous. The main observation is that for any open rectangle
$$h^{-1}(U \times V) = f^{-1}(U) \cap g^{-1}(V)$$
is open in Z (being the intersection of two open sets). In general, any open set in the product topology can be written as $W = \cup_i U_i \times V_i$, and
$$h^{-1}(W) = h^{-1}\big(\bigcup_i U_i \times V_i\big) = \bigcup_i h^{-1}(U_i \times V_i)$$
is open in Z (being the union of open sets).

Conclusion: $X \times Y$ is the categorical product in Top.

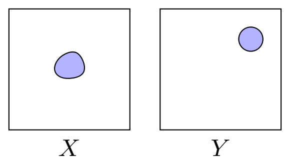
Example 3.23. Consider $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ with the product topology (with usual topology on $\mathbb{R}$). Let

$$f : Z \to \mathbb{R}^2, \qquad f(z) = (f_1(z), f_2(z)).$$

Then f is continuous iff f_1 and f_2 are continuous.

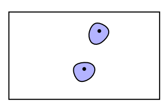
3.4.7. Categorical coproduct. For topological spaces X and Y, their categorical coproduct is given by their disjoint union $X \coprod Y$.

By definition, a subset in the disjoint union is open if its intersection with each piece is open.



There are canonical inclusions $X \to X \coprod Y$ and $Y \to X \coprod Y$ which are both continuous. For any topological space Z and continuous maps $f : X \to Z$ and $g : Y \to Z$, the map $f \coprod g : X \coprod Y \to Z$ is continuous.

3.4.8. Hausdorff topological spaces. A *neighborhood* of a point x in a topological space X is an open set in X containing x. A topological space X is *Hausdorff* if for any distinct points $x, y \in X$, there exist neighborhoods of x and y which are disjoint (that is, their intersection is empty).



This is a property of a topological space. It says that distinct points can be separated by neighborhoods.

Example 3.24. Metric spaces are Hausdorff: Given $x, y \in X$, the balls $B(x, r)$ and $B(y, r)$ do not intersect whenever $0 < r < \frac{1}{2}d(x, y)$. This follows from the triangle inequality.

The indiscrete topology on a set with at least two points is not Hausdorff (since any open set containing a point is necessarily the full space).

Problems.

(1) For $X = \{a, b, c\}$, list all the subsets of 2^X which define a topology on X. For instance, $\tau = \{\emptyset, X, \{a, b\}, \{c\}\}$ is a topology on X. If τ is a topology on X, then is τ^c (obtained by taking complement of each subset in τ) also a topology?

(2) Describe the initial object, terminal object, product, coproduct in the category Top. The product and coproduct are already discussed.

(3) To every topological space X, one can associate a poset consisting of the open sets of X ordered by inclusion. (That is, view τ as a subposet of the Boolean poset 2^X.) Does this give rise to a functor $\mathsf{Top} \to \mathsf{Poset}$?

(4) Show that: The underlying topology of the discrete metric is the discrete topology. If a set X has more than one element, then the indiscrete topology on X is not metrizable.

(5) Show that: Let $f : X \to Y$ be any function between topological spaces. Then f is a homeomorphism iff f is a bijection and induces a bijection between the set of open sets of X and the set of open sets of Y.
(6) Consider two different topologies τ and τ' on the same set X such that τ strictly contains τ'. Show that: The identity map $(X, \tau) \to (X, \tau')$ is continuous but its inverse which is also the identity map is not continuous. This shows that a continuous bijection between topological spaces may not be a homeomorphism.
(7) Show that: The discrete and indiscrete topologies on a set give rise to functors

$$\mathsf{Set} \to \mathsf{Top},$$

and these are the left and right adjoints, respectively, to the forgetful functor from Top to Set.
(8) Show that: The euclidean, diamond, square metrics on $\mathbb{R}^2$ have the same underlying topology. (When we say continuous map from $\mathbb{R}^2$ to $\mathbb{R}$, it is wrt this topology.) Further, check that it coincides with the product topology on $\mathbb{R} \times \mathbb{R}$.
(9) Let (X, d) be a metric space, and Y be a subset of X with the induced metric. Show that: Show that the underlying topology of Y is the same as the subspace topology of Y obtained from the induced topology on X.
(10) Show that: If X and Y are metrizable, then so is their product and coproduct. (As a consequence of Proposition 3.16, the (co)product in Metric can be constructed by metrizing the (co)product of their images in Top.)
(11) Show that any open set in $\mathbb{R}$ is a countable union of open intervals. Generalize this result to $\mathbb{R}^2$.
(12) Suppose Y is an open set in a topological space X. Show that: A subset of Y is open in the subspace topology iff it is open in X.
(13) Show that: If $f : X \to Y$ is continuous, then so is the induced map $g : X \to f(X)$ (with subspace topology on the latter).
(14) Show that: On a metric space (X, d), the standard bounded metric $\bar{d}$ induces the same topology as d.
(15) Let X be a finite topological space. Show that the following are equivalent.
  (a) X has the discrete topology.
  (b) X is metrizable.
  (c) X is Hausdorff.
(16) Give an example of a Hausdorff space which is not metrizable.
(17) Show that: If X and Y are Hausdorff topological spaces, then so is $X \times Y$ (with product topology).
(18) The diagonal of a set X is defined to $\{(x, x) \mid x \in X\}$. It is a subset of $X \times X$. Show that: A topological space X is Hausdorff iff the diagonal of X is a closed subset of $X \times X$ (with product topology).

3.5. Closure operators

We briefly discuss closure operators on a poset. These were introduced by Birkhoff in his book on lattice theory. A detailed treatment can be found in [9, Chapter 7].

Topological spaces can also be formulated using the notion of closed sets, which are complements of open sets. There is a closure operator on the Boolean poset of a topological space whose closed sets are precisely the closed sets of the topology.

3.5.1. Closure operators. Let P be any poset (not necessarily finite). A *closure operator* on P is a map

$$c : P \to P$$

such that for every $x, y \in P$, we have:

- $x \leq c(x)$;
- if $x \leq c(y)$, then $c(x) \leq c(y)$.

We refer to $c(x)$ as the closure of x.

One can deduce that for a closure operator c,

- $c\big(c(x)\big) = c(x)$;
- if $x \leq y$, then $c(x) \leq c(y)$.

Thus, a closure operator is necessarily order-preserving.

An element $z \in P$ is *closed* wrt a closure operator c if $c(z) = z$, or equivalently if $z = c(x)$ for some $x \in P$.

Recall that 2^I denotes the Boolean poset on the set I. A closure operator $t : 2^I \to 2^I$ is *topological* if

- $t(\emptyset) = \emptyset$;
- $t(A \cup B) = t(A) \cup t(B)$.

There is a dual notion of coclosure operator which we do not discuss.

3.5.2. Closed sets in a topological space. Let X be a topological space. A subset of X is *closed* if its complement in X is open. The collection of closed sets in a topological space satisfy the following axioms. The empty set and the full space are closed, arbitrary intersection of closed sets is closed, and finite union of closed sets is closed.

Observe that by interchanging the roles of intersections and unions we go from the axioms for closed sets to the axioms for open sets. The two sets of axioms are distinct since intersection and union play an asymmetric role in the definition of topological spaces.

3.5.3. Closure operator of a topological space. Given any subset A of a topological space X, define the *closure* of A to be the intersection of all closed sets containing A. We denote the closure of A by $\overline{A}$. Clearly, $\overline{A}$ is a closed set, and $A \subseteq \overline{A}$. Also if $A \subseteq \overline{B}$, then $\overline{A} \subseteq \overline{B}$. This shows that

$$2^X \to 2^X, \qquad A \mapsto \overline{A}$$

is a closure operator. Further,

$$\overline{\emptyset} = \emptyset \quad \text{and} \quad \overline{A \cup B} = \overline{A} \cup \overline{B}.$$

So this is a topological closure operator. It is clear that the closed sets of this closure operator are precisely the closed sets of the topology.

Example 3.25. Consider $\mathbb{R}$ with usual topology. The subset $A := \mathbb{N}$ is closed, so its closure is itself. The closure of $A := \{1/n \mid n \in \mathbb{N}_+\}$ is $A \cup \{0\}$. The closure of $A := [0, 1)$ is $[0, 1]$.

We say that a set A *intersects* a set B if $A \cap B \neq \emptyset$.

Lemma 3.26. *Let A be a subset of a topological space X. Then $x \in \overline{A}$ iff every open set containing x intersects A.*

Proof. The following statements are equivalent.

- $x \in \overline{A}$.
- If A is a subset of a closed set Z, then Z contains x.
- If an open set U does not intersect A, then U does not contain x.
- If an open set U contains x, then it intersects A.

□

3.5.4. Interior of a set. Given any subset A of a topological space X, define the *interior* of A to be the union of all open sets contained in A. We denote the interior of A by A°. Clearly, A° is an open set, and $A^\circ \subseteq A$. Also if $A^\circ \subseteq B$, then $A^\circ \subseteq B^\circ$. The order-preserving map

$$2^X \to 2^X, \qquad A \mapsto A^\circ$$

is a coclosure operator (which we have not discussed).

3.5.5. Limit points. A *neighborhood* of a point x in a topological space X is an open set in X containing x. For a subset A of X, we say $x \in X$ is a *limit point* of A if every neighborhood of x intersects A in a point other than x.

Example 3.27. Consider $\mathbb{R}$ with usual topology. The subset $A := \mathbb{N}$ has no limit limits. The subset $A := \{1/n \mid n \in \mathbb{N}_+\}$ has exactly one limit point, namely 0. The set of limit points of $A := (0, 1]$ is $[0, 1]$.

The terms *cluster point* or *accumulation point* are also commonly used instead of limit point.

Lemma 3.28. *Let A be a subset of a topological space X. Let A' denote the set of all limit points of A. Then $\overline{A} = A \cup A'$.*

Proof. By definition, $A \subseteq \overline{A}$ and by Lemma 3.26, $A' \subseteq \overline{A}$. Hence $A \cup A' \subseteq \overline{A}$. For getting equality, let $x \in \overline{A}$. If $x \in A$, then we are done. If not, then any neighborhood of x intersects A by Lemma 3.26 in a point other than x, and hence $x \in A'$. □

3.5.6. Convergence in topological spaces. One can consider sequences in any topological space. A sequence (a_n) in a topological space X *converges* to a point $a \in X$, if for any neighborhood U of a, there exists an integer N such that $a_n \in U$ for all $n > N$.

This definition is sensible. However, standard results of sequences in metric spaces do not carry over in this generality. For instance, a sequence can converge to more than one point [18, pages 98-100]. There are gadgets called *nets* more general than sequences which work better [18, pages 187-188] or [13, Chapter 2].

Problems.

(1) Let X be any set and $t : 2^X \to 2^X$ be a topological closure operator. Show that: The closed sets of this closure operator satisfy the axioms of closed sets for a topology on X.

(2) Let $f : X \to Y$ be any function between topological spaces X and Y. Show that the following are equivalent.

  (a) f is continuous.

(b) For every closed set B in Y, $f^{-1}(B)$ is a closed set in X.
(c) For every subset A of X, $f(\overline{A}) \subseteq \overline{f(A)}$.

(3) Show that: A closed subset of a complete metric space is complete in the induced metric.
(4) Let (X, d) be a metric space, and let A be any subset. Show that $x \in \overline{A}$ iff there is a sequence in A which converges to x.
(5) Let Y be a subset of a topological space X. Show that: A set A is closed in Y (under the subspace topology) iff A is the intersection of some closed set of X with Y.
(6) (Pasting lemma.) Let $X = A \cup B$, where A and B are closed sets in a topological space X. Let Y be any topological space, and $f : X \to Y$ be any function. Show that: f is continuous iff the restrictions $f|_A : A \to Y$ and $f|_B : B \to Y$ are both continuous wrt the subspace topologies on A and B. Generalize this result to a finite number of closed sets whose union is X.
(7) Let X be any set. Call a subset U of X open if either U is empty or $X \setminus U$ is a finite set. Show that this defines a topology τ on X. How does this differ from the discrete topology on X? Describe the closed sets of this topology. Is this topology Hausdorff?
(8) Let A be a nonempty closed subset of $\mathbb{R}$ which is bounded (both above and below). Show that the least upper bound and the greatest lower bound of A both belong to A.
(9) Let Y be the triangle in $\mathbb{R}^2$ joining the points $(0, 0)$, $(0, 1)$, $(1, 0)$. Give it the subspace topology induced from the usual topology on $\mathbb{R}^2$. Similarly, let $X = [0, 1)$ with subspace topology from $\mathbb{R}$. Consider the map $f : X \to Y$ which maps $[0, 1/3]$ to the line-segment joining $(0, 0)$ and $(1, 0)$, $[1/3, 2/3]$ to the line-segment joining $(1, 0)$ and $(0, 1)$, and $[2/3, 1)$ to the line-segment joining $(0, 1)$ and $(0, 0)$ (in a linear manner). Show that: f is a continuous bijection which is not a homeomorphism. In particular, the inverse f^{-1} is not continuous. Give an example of a (open) closed set Z in X such that $f(Z)$ is not (open) closed.
(10) If X is a Hausdorff topological space, then a sequence in X converges to at most one point of X.

3.6. Compactness

We discuss compactness. It is a finiteness property of topological spaces and plays a role similar to finite sets in the category of sets.

3.6.1. Definition and examples. Let X be a topological space. A *cover* of X is a collection $\mathcal{A}$ of subsets of X whose union is X. A subcover of a cover $\mathcal{A}$ is a subset of $\mathcal{A}$ whose union is also X. An *open cover* of X is a cover of X consisting of open sets.

A topological space X is *compact* if every open cover of X has a finite subcover. This is a property of a topological space.

Example 3.29. The real line $\mathbb{R}$ with usual topology is not compact. For instance, the open cover

$$\mathcal{A} = \{(n - 2/3, n + 2/3) \mid n \in \mathbb{Z}\}$$

does not have a finite subcover. (The choice of 2/3 is somewhat arbitrary.)

A finite topological space is compact (since the number of open sets is itself finite). An infinite topological space with discrete topology is not compact (since each point by itself is an open set). To summarize: a discrete topological space is compact iff it is finite.

The subspace topology on $X = \{1/n \mid n \in \mathbb{N}_+\}$ induced from $\mathbb{R}$ is discrete. Hence, X is not compact. However, let us add the limit point of X and see what happens. The subspace topology on $X \cup \{0\}$ is no longer discrete (since $\{0\}$ is not an open set). In fact, any open set containing $\{0\}$ will contain all but finitely many elements of X. Therefore, $X \cup \{0\}$ is compact.

The open interval $(0, 1)$ (with usual topology) is not compact. For instance, the open cover

$$\mathcal{A} = \{(1/n, 1) \mid n \in \mathbb{N}_+\}$$

does not have a finite subcover.

A topological space is either compact or not. So, one can consider the full subcategory of Top whose objects are compact topological spaces. We denote it by $\mathsf{CompactTop}$. A metric space is compact if its underlying topology is compact. So, similarly, one can consider the full subcategory of Metric whose objects are compact metric spaces.

Theorem 3.30. *Any closed interval $[a, b]$ of the real line is compact.*

Proof. We may assume $a < b$, since otherwise the result is clear. Let $\mathcal{A}$ be an open cover of $[a, b]$. We need to find a finite subcover. Consider the set

$$\mathrm{B} = \{r \in [a, b] \mid [a, r] \text{ is a subset of the union of finitely many elements of } \mathcal{A}\}.$$

Then B is nonempty (since $a \in \mathrm{B}$) and bounded above by b. In fact, it is clear that all points in some open interval containing a, and bigger than a, belong to B. Hence by Proposition 2.16, B has a least upper bound. Call it c. Clearly $c > a$. We claim that $c \in \mathrm{B}$ and $c = b$: Suppose $c \neq b$. Pick an element U of $\mathcal{A}$ which contains c. Since U is open, it contains points c_1 and c_2 such that $c_1 < c < c_2$. Since c is the least upper bound, $c_1 \in \mathrm{B}$, and hence $[a, c_1]$ is a subset of the union of finitely many elements of $\mathcal{A}$. By throwing in the element U, we deduce that $[a, c_2]$ is also a subset of the union of finitely many elements of $\mathcal{A}$. Hence $c_2 \in \mathrm{B}$, which is a contradiction. This shows that $c = b$. To show that $c \in \mathrm{B}$, we argue similarly, pick U, and then $c_1 < c$ in U, so $[a, c_1]$ is a subset of the union of finitely many elements of $\mathcal{A}$, and hence so is $[a, c]$. □

Lemma 3.31. *If X is compact and Z is a closed set of X, then Z is compact (in the subspace topology).*

Proof. Let $\{V_i\}_{i \in I}$ be an open cover of Z. Since Z has the subspace topology, $V_i = U_i \cap Z$ for some open set U_i in X. Further, since Z is a closed set, its complement $X \setminus Z$ is an open set. Hence $\{U_i\}_{i \in I} \cup \{X \setminus Z\}$ is an open cover of X. Since X is compact, this has a finite subcover. The corresponding finitely many V_i cover Z. Thus Z is compact. □

Lemma 3.32. *Every compact subspace of a Hausdorff space is closed.*

Proof. Let X be a Hausdorff space and Y be a compact subspace of X. We show that $X \setminus Y$ is open. Let z be any point in U. We need to find a neighborhood of z which is contained in $X \setminus Y$. To that end, for each $y \in Y$, let V_y and U_y be neighborhoods of y and z which do not intersect each other. Now $\{V_y \cap Y\}$ is an open cover of Y. Since Y is compact, it has a finite subcover, say $\{V_{y_1} \cap Y, V_{y_2} \cap Y, \ldots, V_{y_k} \cap Y\}$. Put

$$U := U_{y_1} \cap U_{y_2} \cap \cdots \cap U_{y_k}.$$

Then U is a neighborhood of z which is contained in $X \setminus Y$. □

Lemma 3.33. *Every compact subspace of a metric space is closed and bounded.*

PROOF. Suppose A is a compact subspace of a metric space X. Since X is Hausdorff, A is closed by Lemma 3.32. Now consider the nested open balls $B(x_0, n)$ of radius n centered at some point x_0. They form an open cover of X:

$$X = \bigcup_{n \geq 1} B(x_0, n).$$

Every point x is X is at a finite distance r from x_0, and by the Archimedean property $r < n$ for some integer n, and then x belongs to $B(x_0, n)$. So the intersections of these open balls with A give an open cover of A. Since A is compact, there exists a finite subcover. So A is a subset of $B(x_0, n)$ for some n, and hence bounded. □

3.6.2. Product of compact sets. We now show that the product of compact sets is again compact.

Lemma 3.34 (Tube Lemma). *Consider $X \times Y$, with Y compact. Let $x_0 \in X$. Let W be an open set of $X \times Y$ which contains the slice $\{x_0\} \times Y$, then there is a neighborhood U of x_0 in X such that $U \times Y \subseteq W$.*

The open set $U \times Y$ is called a tube around $\{x_0\} \times Y$.

PROOF. For each $y \in Y$ pick a neighborhood $U_y \times V_y$ of (x_0, y) which is a subset of W. Thus U_y is a neighborhood of x_0 in X and V_y is a neighborhood of y in Y. Then $\{V_y\}_{y \in Y}$ is an open cover of Y. Since Y is compact, it has a finite subcover, say $\{V_{y_1}, V_{y_2}, \ldots, V_{y_k}\}$. Put

$$U := U_{y_1} \cap U_{y_2} \cap \cdots \cap U_{y_k}.$$

Then U is an open set in X, and by construction, $U \times Y \subseteq W$. □

Theorem 3.35. *If X and Y are compact, then so is $X \times Y$.*

PROOF. Suppose $\mathcal{A}$ is an open cover of $X \times Y$. For each $x \in X$, the subspace $\{x\} \times Y$ is compact, so it is contained in the union of finitely many elements of $\mathcal{A}$. By Lemma 3.34, we deduce that for each $x \in X$, there is a neighborhood U_x of x in X such that $U_x \times Y$ is contained in the union of finitely many elements of $\mathcal{A}$. Now $\{U_x\}_{x \in X}$ is an open cover of X. Since X is compact, it has a finite subcover say $\{U_1, \ldots, U_k\}$. Then the tubes $U_i \times Y$ for $1 \leq i \leq k$ cover $X \times Y$, and each tube is contained in the union of finitely many elements of $\mathcal{A}$. This gives a finite subcover of $X \times Y$. □

Given topological spaces $X_1, \ldots, X_k$, one can form their cartesian product

$$X_1 \times \cdots \times X_k.$$

We can give this the product topology: A subset is open if it can be expressed as a union of higher rectangles $U_1 \times \cdots \times U_k$ with U_i open in X_i for each i. For $k = 2$, this recovers the previous construction. Under the product topology, the canonical projections

$$X_1 \times \cdots \times X_k \to X_i$$

are continuous for each i. Further, if Z is any topological space with continuous maps $f_i : Z \to X_i$ for each i, then the induced map

$$Z \to X_1 \times \cdots \times X_k, \qquad z \mapsto (f_1(z), \dots, f_k(z))$$

is continuous. Hence $X_1 \times \cdots \times X_k$ is the categorical product of $X_1, \dots, X_k$. Any nonempty finite product can be constructed iteratively. Explicitly, this implies that the canonical map

$$(X_1 \times \cdots \times X_{k-1}) \times X_k \longrightarrow X_1 \times \cdots \times X_k$$

is a homeomorphism. You may also check this directly.

Theorem 3.36. *If $X_1, \dots, X_k$ are compact, then so is $X_1 \times \cdots \times X_k$.*

Proof. We know that

$$(\dots((X_1 \times X_2) \times X_3) \times \cdots \times X_k) \quad \text{and} \quad X_1 \times \cdots \times X_k$$

are homeomorphic. By repeated application of Theorem 3.35, we deduce that the former is compact. Since compactness is preserved under homeomorphisms, it follows that the latter is also compact. □

It is true that the product of an arbitrary number of compact spaces is compact (with appropriate topology). This is known as Tychonoff's theorem.

3.6.3. Compact subsets of euclidean space. Let $\mathbb{R}^n$ denote the n-fold cartesian product of $\mathbb{R}$ with itself. The euclidean, diamond, square metrics generalize to this setting. For instance, the square metric is defined by

$$d_\infty((x_1, \dots, x_n), (y_1, \dots, y_n)) := \max\{|x_1 - y_1|, \dots, |x_n - y_n|\}$$

It is easy to verify that this is a metric. We refer to its underlying topology as the standard topology on $\mathbb{R}^n$. See also Example 3.17. Since it arises from a metric, it is Hausdorff. Further, it coincides with the product topology on $\mathbb{R}^n$.

Recall the notion of bounded subsets in a metric space. One can check that a subset A of $\mathbb{R}^n$ is bounded iff A is a subset of some open ball centered at 0.

Theorem 3.37. *Let A be a subspace of $\mathbb{R}^n$ (in the standard topology). Then A is compact iff A is closed and bounded.*

Proof. In view of Lemma 3.33, we only need to prove the backward implication. Suppose A is closed and bounded. Since A is bounded, there is a n such that A is a subset of the open ball $B(0, n)$ centered at the origin. (The open ball is wrt the square metric.) The closure of this open ball is a cartesian product of closed intervals. So by Theorems 3.30 and 3.36, it is compact. Since A is a closed subset of a compact set, it is compact by Lemma 3.31. □

This is called the Heine-Borel theorem. It is a very useful characterization of compact sets since being closed and bounded is usually easy to verify in particular examples. For instance, the theorem implies that any closed interval $[a, b]$ is compact. (It is a different matter that we first proved this result and used it to prove the theorem.)

Example 3.38. Consider the subsets $X = \{1/n \mid n \in \mathbb{N}_+\}$ and $X \cup \{0\}$ of $\mathbb{R}$. Both are bounded. But X is not closed. Its closure is precisely $X \cup \{0\}$. Hence, X is not compact while $X \cup \{0\}$ is compact (consistent with what we had observed earlier).

3.6.4. Compactness and continuity.

Lemma 3.39. *Suppose $f : X \to Y$ is a continuous map between topological spaces, and X is compact. Then $f(X)$ is compact (as a subspace of Y under the subspace topology).*

Proof. Let $\mathcal{A}$ be an open cover of $f(X)$. Consider the collection $\mathcal{A}'$ obtained by replacing each element $U \in \mathcal{A}$ by an open set V in Y such that $V \cap f(X) = U$. Thus $\mathcal{A}'$ consists of open sets in Y. The collection

$$\{f^{-1}(V) \mid V \in \mathcal{A}'\}$$

is an open cover of X. So it has a finite subcover, say, $\{f^{-1}(V_1), \dots, f^{-1}(V_k)\}$. Then $\{V_1 \cap f(X), \dots, V_k \cap f(X)\}$ is a finite subcover of $\mathcal{A}$. □

Theorem 3.40. *Suppose X is a nonempty compact space and $f : X \to \mathbb{R}$ is a continuous function. Then there exist points x_0 and x_1 in X such that $f(x_0) \leq f(x) \leq f(x_1)$ for all $x \in X$.*

In other words, a real-valued continuous function on a compact space attains its infimum and supremum values. For $X = [a, b]$, this result specializes to Theorem 2.25. A more general result is given in [18, Theorem 27.4].

Proof. By Lemma 3.39, $f(X)$ is compact (with subspace topology from $\mathbb{R}$). By Theorem 3.37, $f(X)$ is a closed and bounded subset of $\mathbb{R}$. So the least upper bound and greatest lower bound of $f(X)$ (taken in the poset $\mathbb{R}$) exist and in fact belong to $f(X)$ (see an earlier exercise). Hence there exist points x_0 and x_1 in X such that $f(x_0)$ is the lub and $f(x_1)$ is the glb. □

3.6.5. Compactness in terms of closed sets. A collection $\mathcal{B}$ of subsets of X has the *finite intersection property* if for every finite subcollection of $\mathcal{B}$, the intersection of all subsets in the subcollection is nonempty.

Proposition 3.41. *Let X be a topological space. Then X is compact iff for any collection $\mathcal{B}$ of closed sets with the finite intersection property, the intersection $\cap_{Z \in \mathcal{B}} Z$ of all elements of $\mathcal{B}$ is nonempty.*

Proof. The following statements are equivalent.

- X is compact.
- For any collection $\mathcal{A}$ of open sets whose union is X, there is a finite subset of $\mathcal{A}$ whose union is X.
- For any collection $\mathcal{B}$ of closed sets with empty intersection, there is a finite subset of $\mathcal{B}$ with empty intersection.
- For any collection $\mathcal{B}$ of closed sets with the finite intersection property, the intersection $\cap_{Z \in \mathcal{B}} Z$ of all elements of $\mathcal{B}$ is nonempty.

Compare with the proof of Lemma 3.26. □

Corollary 3.42. *A nested sequence $Z_1 \supseteq Z_2 \supseteq \dots$ of nonempty closed sets in a compact space X has a nonempty intersection (that is, there is a point $x \in X$ which belongs to all the Z_i).*

PROOF. The intersection of a finite number of Z_i is one of them. Since the Z_i are assumed to be nonempty, the collection of the Z_i satisfies the finite intersection property. The result now follows from Proposition 3.41. □

3.6.6. Characterizations of compactness. Let X be a topological space. It is *limit point compact* if every infinite subset of X has a limit point. It is *sequentially compact* if every sequence in X has a convergent subsequence.

Example 3.43. Consider $X = (0, 1)$ with the standard metric. The set $\{1/n \mid n \in \mathbb{N}_+\}$ is infinite but it has no limit point. Similarly, the sequence $(1/n)$ does not have any convergent subsequence. So X is not limit point compact, and also not sequentially compact.

Now consider $X = [0, 1]$ with the standard metric. It is both limit point compact and sequentially compact. See if you can prove this. It will be a consequence of a general result that we prove below. Note that the set $\{1/n \mid n \in \mathbb{N}_+\}$ does have a limit point, namely 0.

Lemma 3.44. *If X is compact, then X is limit point compact.*

PROOF. Let A be a subset of X with no limit point. We need to show that A is finite. By hypothesis, each $x \in X$ has a neighborhood U_x which does not intersect A at any point other than x. This has a finite subcover, say $U_{x_1}, \dots, U_{x_k}$. The union of these sets is X and each of them contains at most one point from A. We conclude that A is finite, as required. □

Lemma 3.45. *Let X be a metric space. If X is limit point compact, then it is sequentially compact.*

PROOF. Let (a_n) be a sequence in X. We want to show that it has a convergent subsequence. Suppose the image of this sequence in X is finite. Then it repeats an element infinitely often. The corresponding subsequence is constant and hence converges. So suppose that the image of (a_n) in X is infinite. By hypothesis, it has a limit point. Call it a. Any neighborhood of a must contain infinitely many points from A. Why? Hence, for each n, there is a a_{i_n} in the open ball $B(a, 1/n)$ such that $i_1 < i_2 < \dots$. Clearly, $a_{i_n} \to a$, and this is the desired subsequence. Alternatively: we can pick a distinct points from each $B(a, 1/n)$, and then rearrange them in increasing order of their indices. (Here we use that the rearrangement of any convergent sequence is convergent.) □

Lemma 3.46. *Let X be a metric space. If X is sequentially compact, then it is totally bounded.*

PROOF. Suppose X is not totally bounded. Let $\epsilon > 0$ be such that X cannot be covered by finitely many ϵ-balls. We now construct a sequence (a_n) which has no convergent subsequence. Pick a_0 to be any point of X, then pick a_1 to be outside $B(a_0, \epsilon)$, and inductively pick a_n to be outside

$$B(a_0, \epsilon) \cup \dots \cup B(a_{n-1}, \epsilon).$$

By construction, the pairwise distance between the a_i is greater than ϵ, so it cannot have a convergent subsequence. □

Lemma 3.47. *Suppose X is a sequentially compact metric space. Then: For any open cover $\mathcal{A}$ of X, there is a real number $\lambda > 0$ such that every ball of radius λ is contained in some open set belonging to $\mathcal{A}$.*

Proof. Let X be sequentially compact. Suppose there is no such λ. Then for each n, there exists a point $a_n \in X$ such that $B(a_n, 1/n)$ is not contained in some open set belonging to $\mathcal{A}$. By sequential compactness, the sequence (a_n) has a convergent subsequence. Say $a_{i_n} \to a$. Now $a \in U$ for some open set $U \in \mathcal{A}$. So for sufficiently small r, $B(a, r) \subseteq U$. Hence for sufficiently large n, $B(a_{i_n}, 1/i_n)$ will be contained in $B(a, r)$, and hence in U. This is a contradiction. □

This is called the Lebesgue number lemma. The positive number λ is called a Lebesgue number of the cover. If λ is a Lebesgue number for a cover, then so are all positive numbers smaller than λ.

The concept of a Lebesgue number makes sense for any open cover of any metric space (not necessarily sequentially compact). It will not exist in general.

Theorem 3.48. *Let X be a metric space. The following are equivalent.*

(1) *X is compact.*
(2) *X is limit point compact.*
(3) *X is sequentially compact.*
(4) *X is complete and totally bounded.*

Proof. (1) implies (2). This was done in Lemma 3.44.

(2) implies (3). This was done in Lemma 3.45.

(3) implies (4). Let X be sequentially compact. Suppose (a_n) is a Cauchy sequence in X. By hypothesis, it has a subsequence which converges to say $a \in X$. Cauchyness forces $a_n \to a$. Thus, A is complete. Lemma 3.46 implies that it is totally bounded.

(3) implies (1). Let X be sequentially compact. Let $\mathcal{A}$ be an open cover of X. We want to find a finite subcover. Let $\lambda > 0$ be as given by Lemma 3.47. Since X is totally bounded, we can cover X by finitely many λ-balls. Each of these balls is contained in some open set belonging to $\mathcal{A}$. These open sets form a finite subcover.

(4) implies (3). Let X be complete and totally bounded. Let (a_n) be a sequence in X. We want to show that it has a convergent subsequence. Since X is complete, we only need to find a Cauchy subsequence. First cover X by finitely many open balls of radius 1. Let U_1 be one of these balls which contains a_n for infinitely many n. That is, some subsequence (a^1_n) of (a_n) lies entirely in U_1. Now cover X by finitely many open balls of radius $1/2$. Let U_2 be one of these balls such that some subsequence (a^2_n) of (a^1_n) lies entirely in U_2. Continue this process. The diagonal (a^n_n) is a Cauchy subsequence. □

In view of Theorem 3.48, the Lebesgue number lemma also holds with sequentially compact replaced by compact. See [18, Lemma 27.5].

Problems.

(1) Let X be compact and Y be Hausdorff. Show that: Any continuous bijection $f : X \to Y$ is a homeomorphism.
(2) Show that CompactTop is an isomorphism-closed subcategory of Top. In other words, if X is compact, and X and Y are homeomorphic, then Y is also compact. (This is often phrased as: Compactness is a topological property.)
(3) Theorem 3.35 implies that the categorical product exists in CompactTop. What about the coproduct?

(4) Let X and Y be metric spaces with X compact. Show that: Any continuous function $f : X \to Y$ is uniformly continuous.
(5) Using first principles, show that: If X is compact, and $f : X \twoheadrightarrow Y$ is a continuous surjective map between topological spaces, then Y is compact. Combine this result with an exercise in Section 3.4 to deduce Lemma 3.39.
(6) Suppose $Z_1 \supseteq Z_2 \supseteq \dots$ is a nested sequence of nonempty closed sets in a compact metric space X. Show that: If the diameters of the Z_n go to zero with n, then $\cap_i Z_i$ is a single point.
(7) Using the definition of sequential compactness, show that: A finite set (with any topology) is sequentially compact.
(8) Give an example of a topological space which is limit point compact but not compact. In view of Theorem 3.48, such a space must necessarily be non-metrizable.
(9) Let X and Y be metric spaces, and let $f : X \to Y$ be any map. Define the graph of f to be

$$G = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y.$$

Show that: G is compact iff X is compact and f is continuous.

3.7. Connectedness

We study connectedness. Roughly speaking, a topological space is connected if it cannot be written as a disjoint union of two nonempty topological spaces.

3.7.1. Connectedness. Let X be a topological space. A *separation* of X is a pair (U, V) where U and V are nonempty disjoint open sets whose union is X. The space X is *connected* if there does not exist a separation of X.

Thus: X is connected iff the only subsets of X which are both open and closed are $\emptyset$ and X.

A topological space is either connected or not. So, one can consider the full subcategory of Top whose objects are connected topological spaces. We denote it by $\mathsf{ConnectedTop}$. A metric space is connected if its underlying topology is connected. So, similarly, one can consider the full subcategory of Metric whose objects are connected metric spaces.

3.7.2. Connected subspaces of $\mathbb{R}$. A subset A of $\mathbb{R}$ is an *interval* if for any $a, b \in A$, $[a, b]$ is contained in A. Note that: open intervals, closed intervals are intervals.

Theorem 3.49. *Let A be any subspace of $\mathbb{R}$ (in the usual topology). Then A is connected iff A is an interval.*

In particular, $\mathbb{R}$ is connected (and so are all open and closed intervals in it).

Proof. Forward implication. Suppose A is not an interval. Then there exist $a, b \in A$ and $c \notin A$ with $a < c < b$. Then

$$(A \cap (-\infty, c), A \cap (c, \infty))$$

is a separation for A. Hence A is not connected.

Backward implication. Suppose A is an interval which is not connected. Let (U, V) be a separation of A. Fix any points $a \in U$ and $b \in V$. Put $U' := U \cap [a, b]$ and $V' := V \cap [a, b]$. Then (U', V') is a separation of $[a, b]$. Let c be the lub of U'.

(It exists because U' is nonempty and bounded above.) Either $c \in U'$ or $c \in V'$. But both lead to a contradiction. Why? □

3.7.3. Product of connected sets.

Lemma 3.50. *Let X be a topological space, and $x_0 \in X$. Suppose for each $i \in I$, we are given a connected subspace A_i of X which contains x_0. Then $\cup_{i \in I} A_i$ is also a connected subspace of X.*

Proof. Put $B := \cup_{i \in I} A_i$. Suppose (U, V) is a separation of B. Say $x_0 \in U$. Then for each $i \in I$, consider $(U \cap A_i, V \cap A_i)$. Since A_i is connected, we have $V \cap A_i = \emptyset$, and hence $A_i \subseteq U$. Thus U contains B contradicting the fact that V is nonempty. □

Theorem 3.51. *If X and Y are connected, then so is $X \times Y$.*

Proof. We may assume that both X and Y are nonempty. Fix a point (x_0, y_0) in $X \times Y$. For each $z \in Y$, define

$$A_z := \{(x, z) \mid x \in X\} \cup \{(x_0, y) \mid y \in Y\}.$$

Both the above subsets of A_z are connected (being homeomorphic to X and Y, respectively). They have the point (x_0, z) in common. Thus each A_z is connected by Lemma 3.50. Now the point (x_0, y_0) is common to all the A_z. So the union of all the A_z, which is X, is connected again by Lemma 3.50. □

We can deduce from here that the product of finitely many connected topological spaces is also connected. For instance, $\mathbb{R}^n$ is connected in the standard topology.

3.7.4. Connected sets and continuity.

Lemma 3.52. *Suppose $f : X \to Y$ is a continuous map between topological spaces, and X is connected. Then $f(X)$ is connected (as a subspace of Y under the subspace topology).*

Proof. Since $f : X \to Y$ is continuous, so is $f : X \to f(X)$. Hence we may assume that f is surjective. We want to show that Y is connected. Suppose not. Then there exists a separation (U, V) of Y. Then $f^{-1}(U)$ and $f^{-1}(V)$ are both nonempty and open, and their union is X. Hence $(f^{-1}(U), f^{-1}(V))$ is a separation of X. This is a contradiction. □

Theorem 3.53. *Suppose f is a continuous real-valued function on a connected space X, and $x, y \in X$. Then for any α lying between $f(x)$ and $f(y)$, there exists a $c \in X$ such that $f(c) = \alpha$.*

Proof. Suppose $f : X \to \mathbb{R}$ and X is connected. Further, suppose $f(x) < \alpha < f(y)$, and α is not in the image of f. Then

$$X = f^{-1}(-\infty, \alpha) \cup f^{-1}(\alpha, \infty),$$

and $(f^{-1}(-\infty, \alpha), f^{-1}(\alpha, \infty))$ is a separation of X. This is a contradiction.

Alternatively: Applying Lemma 3.52, we note that $f(X)$ is connected. If α were not in the image of f, then

$$(f(x) \cap (-\infty, \alpha), f(x) \cap (\alpha, \infty)).$$

is a separation of $f(X)$, which is a contradiction.

Alternatively: Applying Lemma 3.52, we note that $f(X)$ is connected. So by Lemma 3.49, $f(X)$ is an interval. Since α lies between $f(x)$ and $f(y)$ and $f(X)$ is an interval, $\alpha \in f(X)$. □

For $X = [a, b]$, this result specializes to Theorem 2.26.

Problems.

(1) Show that $\mathsf{ConnectedTop}$ is an isomorphism-closed subcategory of Top. In other words, if X is compact, and X and Y are homeomorphic, then Y is also compact. (This is often phrased as: Connectedness is a topological property.)
(2) Show that the spaces $(0, 1)$ and $(0, 1]$ (with topologies induced from the standard topology of $\mathbb{R}$) are not homeomorphic.
(3) Show that $\mathbb{R}^2 \setminus \{0\}$ is connected.
(4) A topological space X is *totally disconnected* if the only connected subspaces of X are the empty set and the singleton sets. Show that:
   - Any discrete topological space is totally disconnected.
   - $\mathbb{Q}$ (with subspace topology from $\mathbb{R}$) is totally disconnected.
(5) Let $f : X \to Y$ be a function between metric spaces. Let G denote the graph of f. Show that: X is connected and f is continuous implies that G is connected. Conversely, if G is connected then X is connected (but f may not be continuous).
(6) Does the coproduct exist in $\mathsf{ConnectedTop}$?

3.8. Normed linear spaces

We briefly discuss normed linear spaces. Roughly, these combine features of a vector space and a metric space.

Let $\Bbbk$ be either $\mathbb{R}$ or $\mathbb{C}$.

3.8.1. Normed linear spaces.

A *normed linear space* is a vector space X over $\Bbbk$ equipped with a function

$$\| \ \| : X \to \mathbb{R}$$

such that for all $x, y \in X$ and $\alpha \in \Bbbk$,

$$\|0\| = 0 \text{ and } \|x\| > 0 \text{ if } x \neq 0,$$

$$\|x + y\| \leq \|x\| + \|y\|,$$

$$\|\alpha x\| = |\alpha| \|x\|.$$

The function $\| \ \|$ is called a *norm* on X. Thus, a normed linear space is a pair $(X, \| \ \|)$, where X is a vector space and $\| \ \|$ is a norm on X. We interpret $\|x\|$ as the length of the vector x.

Note: X could be infinite-dimensional.

Example 3.54. The set of real numbers $\mathbb{R}$ with $\|x\| := |x|$ is a normed linear space, see (2.4).

Linear subspaces of normed spaces are again normed spaces.

There are many interesting choices for a norm on $\mathbb{R}^2$: The euclidean norm

$$\|(x_1, x_2)\|_2 := \sqrt{|x_1|^2 + |x_2|^2},$$

the diamond norm

$$\|(x_1, x_2)\|_1 := |x_1| + |x_2|,$$

the square norm

$$\|(x_1, x_2)\|_\infty := \max\{|x_1|, |x_2|\}$$

are three of the most commonly considered norms. Check that these are indeed norms. (For these norms, the set of points of norm 1 form a circle, diamond and square respectively.) These definitions generalize in an obvious manner to $\mathbb{R}^n$, the n-fold cartesian product of $\mathbb{R}$. There is an even more general context for these norms where they are called l^2, l^1, l^∞ respectively.

Let X and Y be normed linear spaces. A *bounded linear transformation* from X to Y is a linear map $f : X \to Y$ for which there exists an $K > 0$ such that

$$\|f(x)\| \leq K\|x\|$$

for all $x \in X$.

One can readily check that the composite of two bounded linear transformations is again a bounded linear transformation. This yields the category NLS whose objects are normed linear spaces, and morphisms are bounded linear transformations.

3.8.2. Functor to metric spaces. Let $(X, \| \ \|)$ be any normed linear space. The function

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

defines a metric d on X. (Check that the metric axioms hold.)

Now suppose $f : X \to Y$ is a bounded linear transformation. Then

$$d(f(x), f(y)) = \|f(x) - f(y)\| = \|f(x - y)\| \leq K\|x - y\| = Kd(x, y).$$

Thus f viewed as a function between the corresponding metric spaces is Lipshitz continuous. This gives a functor

$$\mathsf{NLS} \to \mathsf{Metric}_L.$$

In view of this, properties of metric or topological spaces also apply to normed linear spaces. For instance, we could ask whether a given normed linear space is complete, and so on.

Note: A series and its convergence makes sense in any normed linear space just as a sequence and its convergence makes sense in any metric space.

Problems.

(1) Let $f : X \to Y$ be a linear map between normed linear spaces. Show that: f is a bounded linear transformation iff f is continuous at 0 (wrt the corresponding metrics).

(2) Let X be a normed linear space. Show that: The norm map is Lipschitz continuous (with usual metric on $\mathbb{R}$).

Bibliography

[1] Tom M. Apostol, *Mathematical analysis*, second ed., Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1974. iv

[2] Mark Anthony Armstrong, *Basic topology*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1983, Corrected reprint of the 1979 original. v

[3] Michael Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991. iv, 14

[4] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969. 9

[5] Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. v

[6] William M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, second ed., Pure and Applied Mathematics, vol. 120, Academic Press Inc., Orlando, FL, 1986. v

[7] Theodor Bröcker and Klaus Jänich, *Introduction to differential topology*, Cambridge University Press, Cambridge, 1982, Translated from the German by C. B. Thomas and M. J. Thomas. v

[8] Andrew Browder, *Mathematical analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1996, An introduction. iv

[9] B. A. Davey and H. A. Priestley, *Introduction to lattices and order*, second ed., Cambridge University Press, New York, 2002. iv, 51

[10] David S. Dummit and Richard M. Foote, *Abstract algebra*, third ed., John Wiley & Sons Inc., Hoboken, NJ, 2004. iv, 14

[11] Paul R. Halmos, *Naive set theory*, Springer-Verlag, New York, 1974, Reprint of the 1960 edition, Undergraduate Texts in Mathematics. iv

[12] Klaus Jänich, *Topology*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1984, With a chapter by Theodor Bröcker, Translated from the German by Silvio Levy. v

[13] John L. Kelley, *General topology*, Springer-Verlag, New York, 1975, Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27. 53

[14] Jeffrey M. Lee, *Manifolds and differential geometry*, Graduate Studies in Mathematics, vol. 107, American Mathematical Society, Providence, RI, 2009. v

[15] Balmohan V. Limaye, *Functional analysis*, second ed., New Age International Publishers Limited, New Delhi, 1996. v

[16] Saunders Mac Lane, *Categories for the working mathematician*, second ed., Graduate Texts in Mathematics, vol. 5, Springer-Verlag, New York, 1998. iv

[17] Shigeyuki Morita, *Geometry of differential forms*, Translations of Mathematical Monographs, vol. 201, American Mathematical Society, Providence, RI, 2001, Translated from the two-volume Japanese original (1997, 1998) by Teruko Nagase and Katsumi Nomizu, Iwanami Series in Modern Mathematics. v

[18] James R. Munkres, *Topology*, Prentice Hall, Inc., Upper Saddle River, NJ, 2000, Second edition of [MR0464128]. iv, v, 42, 53, 58, 60

[19] Charles Chapman Pugh, *Real mathematical analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2002. iv, 29

[20] H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. v

[21] Walter Rudin, *Principles of mathematical analysis*, third ed., McGraw-Hill Book Co., New York, 1976, International Series in Pure and Applied Mathematics. iv, v, 17, 26, 27, 28, 29

[22] George F. Simmons, *Introduction to topology and modern analysis*, Robert E. Krieger Publishing Co. Inc., Melbourne, Fla., 1983, Reprint of the 1963 original. iv

[23] Michael Spivak, *A comprehensive introduction to differential geometry. Vol. I-V*, second ed., Publish or Perish Inc., Wilmington, Del., 1979. v

[24] Terence Tao, *Analysis. I*, second ed., Texts and Readings in Mathematics, vol. 37, Hindustan Book Agency, New Delhi, 2009. iv

[25] ______, *Analysis. II*, second ed., Texts and Readings in Mathematics, vol. 38, Hindustan Book Agency, New Delhi, 2009. iv

[26] Frank W. Warner, *Foundations of differentiable manifolds and Lie groups*, Graduate Texts in Mathematics, vol. 94, Springer-Verlag, New York, 1983, Corrected reprint of the 1971 edition. v