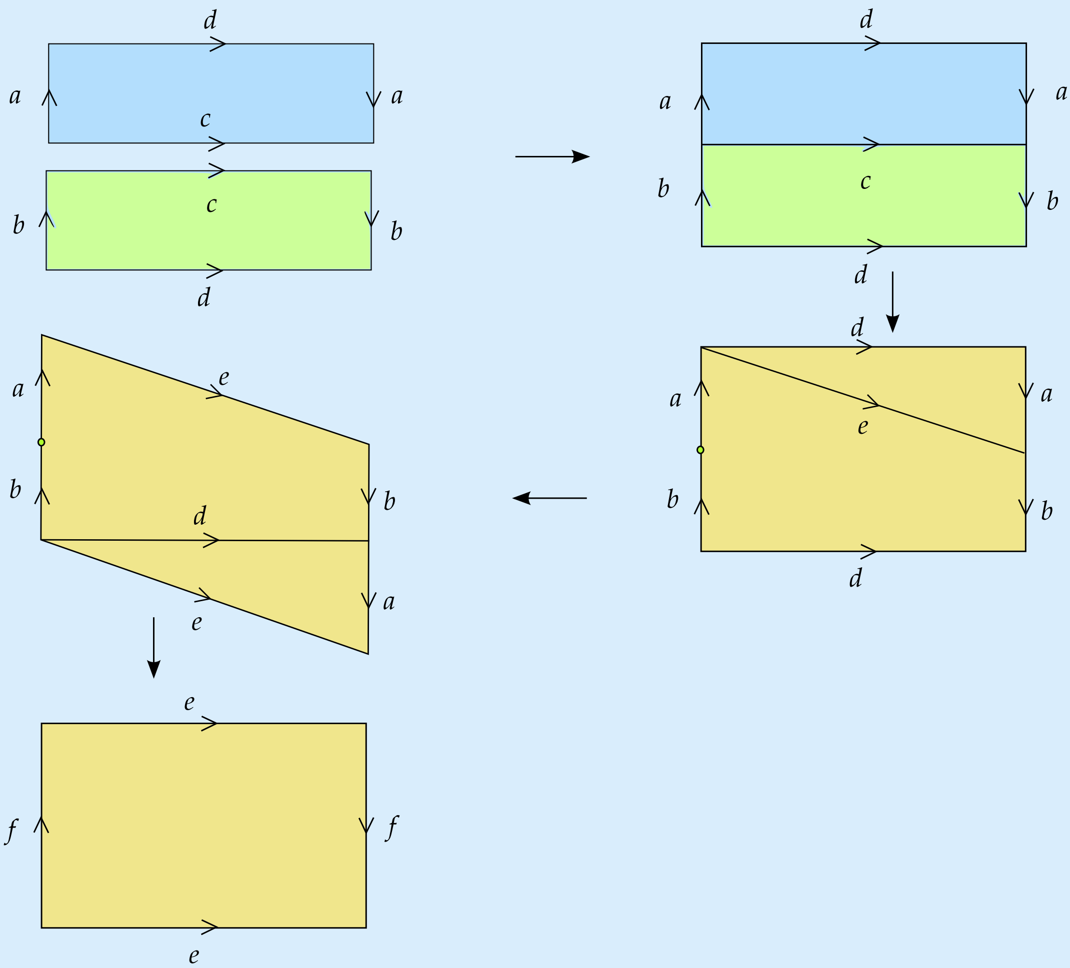
TOPOLOGY

Lectures on



π_1

Huỳnh Quang Vũ

Lecture notes on Topology

Huỳnh Quang Vũ

Version of September 15, 2025

This is a set of lecture notes for courses in topology for undergraduate students at the University of Science in Ho Chi Minh City since 2006. Parts of these notes have also been used in courses for master students at the University of Da Lat and at the University of Education in Ho Chi Minh City. These notes have been written for and tailored to students in those classes.

A sign ✓ in front of a problem notifies that this problem deserves more attention from the reader as it can be used later in the text. A sign * indicates a relatively more difficult topics. Some problems have suggestions placed near the end of the document.

Please send comments to the following address: Huỳnh Quang Vũ, Faculty of Mathematics and Computer Science, University of Science Vietnam National University Ho Chi Minh City. Email: hqvu@hcmus.edu.vn

The latest version of this set of notes together with source files are available at
https://sites.google.com/view/hqvu/teaching

Contents

Introduction

Topology is a mathematical subject that studies shapes. The term comes from the Greek words "topos" (place) and "ology" (study). A set becomes a topological space when each element of the set is given a collection of neighborhoods. Operations on topological spaces must be continuous, bringing certain neighborhoods into neighborhoods. There is no notion of distance. Topology is a part of geometry that does not concern distance.

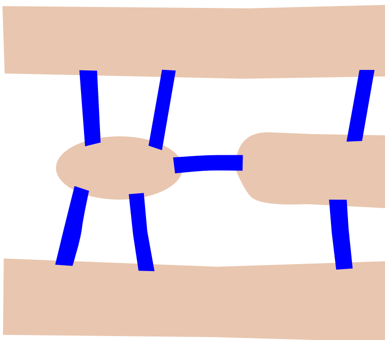


Figure 0.1: The problem "Seven bridges of Konigsberg", studied by Leonard Euler in the 18th century. How to make a closed trip such that every bridge is crossed exactly once? It does not depend on the sizes of the bridges.

Characteristics of topology

Operations on topological objects are more relaxed than in geometry: beside moving around (allowed in geometry), stretching or bending are allowed in topology (not allowed in geometry). For example, in topology circles - big or small, anywhere - are same. Ellipses and circles are same. On the other hand in topology tearing or breaking are not allowed: circles are still different from lines. While topological operations are more flexible they still retain some essential properties of spaces.

Contributions of topology

Topology provides basic notions to areas of mathematics where there is a need for a notion of continuity. It can be useful where metrics or coordinates are not available, not natural, or not necessary.

Initially developed in the late nineteenth century and early twentieth century to provide basis for abstract mathematical analysis, topology gradually became an influential subject, reaching many achievements in the mid and late twentieth century.

Notions and results on continuity is now used throughout mathematics. The ideas of homology is now used widely in geometry and algebra. The concept of manifold has become fundamental in geometry and physics, and is increasingly seen in applications.

Topology often does not stand alone: nowadays there are fields such as algebraic topology, differential topology, geometric topology, combinatorial topology, quantum topology, … Topology often does not solve a problem by itself, but contributes important understanding, settings, and tools. Topology features prominently in differential geometry, global analysis, algebraic geometry, theoretical physics, …. Topology is appearing in contemporary research in applied fields such as computation and data analysis.

Benefits of studying topology

Emerging later than major branches of mathematics such as geometry, algebra, and analysis, it could be said that topology contains concepts, methods, and ideas which are significant advances in mathematical thinking. Thus it is very valuable for students of mathematics to study topology to develop ability to think in general, abstract, and also concrete manners.

The widespread use of topology implies that studying topology can open students to much wider choices of areas for subsequent works and careers.

Part I General Topology

General Topology, also called Point-set Topology, studies fundamentals of neighborhoods, limits, and continuity – basic notions used throughout mathematics. This subject is beneficial and necessary for areas of mathematics which employ notions of continuity, first of all Mathematical Analysis, Geometry, then Algebra, Optimization, Mathematical Physics, and cross-disciplinary areas, and more recently Applied Mathematics.

The course is also for training of mathematical reasoning: rigor, generalization, abstraction.

The course is primarily intended for students from third year, having passed the course in Functional Analysis.

Many arguments and proofs in this part, despise abstract looks, are actually quite straight forward in ideas and simple in techniques. For more effective learning, it is recommended that students try to prove statements themselves, only occasionally read parts of the proofs for suggestions.

There are many textbooks and lecture notes for General Topology, students can use whichever they find readable. [Mun00] is contemporary, containing many examples, figures, and exercises. [Kel55] is classical and is more abstract. [HY61] contains more advanced topics. [AF08] is more elementary and contains applications.

1 Infinite sets

In General Topology we often work in very general settings, in particular we often deal with infinite sets, therefore we start with a study of infinite sets.

We work with "naive set theory", pioneered by Georg Cantor in the late 19th century. We do not define what a set is, thus "set" is an undefined notion. We use basic notions such as elements of a set, the empty set, subsets, unions and intersections of sets, ... We assume the existence of the set of natural numbers, and from it the set of integer numbers, the set of rational numbers, and the set of real numbers together with their properties familiar to us in previous courses.

We do not distinguish set, **class** [1], or **collection** [2], but in occasions where we deal with "set of sets", following common convention, we often prefer the term collection.

Relation

A **relation** [3] R on a set S is a non-empty subset of the set $S \times S$.

When $(a, b) \in R$ we often say that a is related to b and often write $a \sim_R b$.

A relation said to be:

(a) **reflexive** [4] if $\forall a \in S, (a, a) \in R$.

(b) **symmetric** [5] if $\forall a, b \in S, (a, b) \in R \Rightarrow (b, a) \in R$.

(c) **anti-symmetric** [6] if $\forall a, b \in S, ((a, b) \in R \land (b, a) \in R) \Rightarrow a = b$.

(d) **transitive** [7] if $\forall a, b, c \in S, ((a, b) \in R \land (b, c) \in R) \Rightarrow (a, c) \in R$.

An **equivalence relation** on S is a relation that is reflexive, symmetric and transitive.

If R is an equivalence relation on S then an **equivalence class** [8] represented by $a \in S$ is the subset $[a] = \{b \in S \mid (a, b) \in R\}$. Two equivalence classes are either coincident or disjoint. The set S is partitioned [9] into the disjoint union of its equivalence classes.

[1] lớp
[2] bộ
[3] quan hệ
[4] phản xạ
[5] đối xứng
[6] phản đối xứng
[7] bắc cầu
[8] lớp tương đương
[9] phân hoạch

Order

An **order** [1] on a set S is a relation R on S that is reflexive, anti-symmetric and transitive.

Note that two arbitrary elements a and b do not need to be comparable; that is, the pair (a, b) may not belong to R. For this reason an order is often called a partial order.

When $(a, b) \in R$ we often write $a \leq b$. When $a \leq b$ and $a \neq b$ we write $a < b$.

If any two elements of S are related then the order is called a **total order** [2] and $(S, \leq)$ is called a **totally ordered set**.

Example. The set $\mathbb{R}$ of all real numbers with the usual order $\leq$ is totally ordered.

Example. Given a set S the **collection of all subsets** of S is denoted by $\mathcal{P}(S)$ or 2^S (see Problem 1.17 for a reason for this notation). With the inclusion relation, $(2^S, \subseteq)$ is a partially ordered set, but is not a totally ordered set if S has more than one element.

Example. Let $(S_1, \leq_1)$ and $(S_2, \leq_2)$ be two ordered sets. The following is an order on $S_1 \times S_2$: $(a_1, b_1) \leq (a_2, b_2)$ if $(a_1 < a_2)$ or $((a_1 = a_2) \wedge (b_1 \leq b_2))$. This is called the **dictionary order** [3].

In an ordered set, the **smallest element** [4] is the element that is smaller than all other elements. More concisely, if S is an ordered set, the smallest element of S is an element $a \in S$ such that $\forall b \in S, a \leq b$. The smallest element of S if exists is unique, denoted by $\min S$.

A **minimal element** [5] is an element which no element is smaller than. More concisely, a minimal element of S is an element $a \in S$ such that $\forall b \in S, b \leq a \Rightarrow b = a$. There can be more than one minimal element.

A **lower bound** [6] of a subset of an ordered set is an element of the set that is smaller than or equal to any element of the subset. More concisely, if $A \subset S$ then a lower bound of A in S is an element $a \in S$ such that $\forall b \in A, a \leq b$.

The definitions of largest element, maximal element, and upper bound are similar.

[1] thứ tự
[2] thứ tự toàn phần
[3] thứ tự từ điển
[4] phần tử nhỏ nhất
[5] phần tử cực tiểu
[6] chặn dưới

Set equivalence

Two sets are said to be **set-equivalent** if there is a bijection [1] from one set to the other set. Clearly being set-equivalent is transitive.

A set is said to be **finite** if it is either empty or is set-equivalent to a subset $\{1, 2, 3, \ldots, n\}$ of the set all positive integers $\mathbb{Z}^+$ for some $n \in \mathbb{Z}^+$. More precisely, a set S is finite if $S = \emptyset$ or there exists $n \in \mathbb{Z}^+$ such that S is set-equivalent to the set $\{i \in \mathbb{Z}^+ \mid i \leq n\}$, in which case n is called **the number of elements** of S, denoted by $|S|$.

If a set is not finite we say that it is **infinite**.

A set is called **countably infinite** [2] if it is equivalent to $\mathbb{Z}^+$. A set is called **countable** if it is either finite or countably infinite.

Example. The set of all natural numbers $\mathbb{N}$ is countable.

Example. The set $\mathbb{Z}$ of all integer numbers is countable. We can count alternatively the positive and the negative integers. We can work out a formula for the counting, such as

$$
\begin{aligned}
&\mathbb{Z} \to \mathbb{Z}^+ \\
m \mapsto n = &\begin{cases} 2m, & m > 0, \\ -2m + 1, & m \leq 0 \end{cases}
\end{aligned}
$$

or the inverse map

$$
\begin{aligned}
&\mathbb{Z}^+ \to \mathbb{Z} \\
n \mapsto m = &\begin{cases} \frac{n}{2}, & n \text{ is even,} \\ \frac{1-n}{2}, & n \text{ is odd.} \end{cases}
\end{aligned}
$$

1.1 Proposition. *Any subset of a countable set is countable.*

Proof. The statement is equivalent to the statement that any subset of $\mathbb{Z}^+$ is countable. Suppose that A is an infinite subset of $\mathbb{Z}^+$, let us proceed to count A. Let $a_1 = \min A$, whose existence is due to the **well-ordering property** [3] of $\mathbb{Z}^+$: every non-empty subset of $\mathbb{Z}^+$ has a smallest element. For each $n > 1$, let $a_n = \min A \setminus \{a_1, a_2, \ldots, a_{n-1}\}$. Since $a_{n-1} < a_n$ the sequence $(a_n)_{n \in \mathbb{Z}^+}$ consists of distinct elements.

Now we show that any element m of A is a_{n_0} for some n_0, thus the sequence $(a_n)_{n \in \mathbb{Z}^+}$ indeed enumerate A. Let $B = \{a_n \mid n \in \mathbb{Z}^+, a_n \geq m\}$, and let $a_{n_0} = \min B$ (notice that $B \neq \emptyset$). Then $a_{n_0} \geq m$. Since $a_{n_0 - 1} < a_{n_0}$ we

[1] song ánh
[2] vô hạn đếm được
[3] tính được sắp tốt

have $a_{n_0-1} \notin B$, thus $a_{n_0-1} < m$. This implies $m \in A \setminus \{a_1, a_2, \dots, a_{n_0-1}\}$, then $a_{n_0} = \min\left(A \setminus \{a_1, a_2, \dots, a_{n_0-1}\}\right)$ implies $a_{n_0} \le m$. Thus $a_{n_0} = m$. □

Example. In order to follow the above proof more easily, we may try to count the set of prime numbers.

1.2 Proposition. *A non-empty set S is countable if and only if there is an injective map from S to $\mathbb{Z}^+$.*

Thus for a set being a countable is being assigned distinct positive integers to the elements.

Proof. If S is countable then immediately from definitions there is an injective map from S to $\mathbb{Z}^+$.

Conversely, suppose that there is an injective map $f : S \to \mathbb{Z}^+$. Then S and $f(S)$ are set-equivalent, while $f(S)$ is countable by 1.1, hence S is countable. □

1.3 Proposition. *A non-empty set S is countable if and only if there is a surjective map from $\mathbb{Z}^+$ to S.*

Thus being countable is to be indexed (with possible duplications) by the set of postive integers, that is, the elements can be written as a sequence $a_1, a_2, a_3, \dots$.

Proof. If S is finite, $S = \{a_1, a_2, \dots, a_n\}$, then let $\phi : \mathbb{Z}^+ \to S$, $\phi(i) = a_i$ for $1 \le i \le n$, and $\phi(i) = a_1$ for $i > n$. If S is countably infinite, then S has a bijection with $\mathbb{Z}^+$.

Conversely, suppose that there is a surjective map $\phi : \mathbb{Z}^+ \to S$. For each $s \in S$ the set $\phi^{-1}(s)$ is non-empty. Let $n_s = \min \phi^{-1}(s)$. The map $s \mapsto n_s$ is an injective map from S to a subset of $\mathbb{Z}^+$, therefore S is countable by 1.1. □

Proposition. *$\mathbb{Z}^+ \times \mathbb{Z}^+$ is countable.*

Proof. We can enumerate $\mathbb{Z}^+ \times \mathbb{Z}^+$ by counting along diagonals, such as by the method shown in the following diagram:

$$\begin{tikzcd}
(1,1) \arrow[r] & (1,2) \arrow[dl] & (1,3) \arrow[dl] & (1,4) \arrow[dl] \\
(2,1) \arrow[urr] & (2,2) \arrow[dl] & (2,3) \arrow[dl] & \\
(3,1) \arrow[uurrr] & (3,2) \arrow[dl] & \dots & \\
(4,1) & & &
\end{tikzcd}$$

For detail we can derive the explicit formula for the counting:

$$\begin{aligned}
&\mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+ \\
&(m,n) \mapsto (1+2+\cdots+((m+n-1)-1)) + m = \frac{(m+n-2)(m+n-1)}{2} + m.
\end{aligned}$$

We check that this map is injective. Let $k = m+n$. Suppose that $\frac{(k-2)(k-1)}{2} + m = \frac{(k'-2)(k'-1)}{2} + m'$. If $k = k'$ then the equation certainly leads to $m = m'$ and $n = n'$. If $k < k'$ then

$$\begin{aligned}
\frac{(k-2)(k-1)}{2} + m &\leq \frac{(k-2)(k-1)}{2} + (k-1) = \frac{(k-1)k}{2} < \\
&< \frac{(k-1)k}{2} + 1 \leq \frac{(k'-2)(k'-1)}{2} + m',
\end{aligned}$$

a contradiction. □

1.4 Proposition. *The union of a countable collection of countable sets is a countable set.*

Proof. By 1.3 the collection can be indexed as $A_1, A_2, \ldots, A_i, \ldots$, and the elements of each set A_i can be indexed as $a_{i,1}, a_{i,2}, \ldots, a_{i,j}, \ldots$. This means there is a surjective map from the index set $\mathbb{Z}^+ \times \mathbb{Z}^+$ to the union $\bigcup_{i\in I} A_i$ by $(i,j) \mapsto a_{i,j}$, hence $\bigcup_{i\in I} A_i$ is countable, again by 1.3. □

Proposition. *The set $\mathbb{Q}$ of all rational numbers is countable.*

Proof. One way is to write $\mathbb{Q} = \bigcup_{q\in\mathbb{Z}^+} \left\{\frac{p}{q} \mid p \in \mathbb{Z}\right\}$, then use 1.4.

Another way is to write each rational number in the form $\frac{p}{q}$ with $q > 0$ and $\gcd(p,q) = 1$, then observe that the map $\frac{p}{q} \mapsto (p,q)$ from $\mathbb{Q}$ to $\mathbb{Z} \times \mathbb{Z}$ is injective, and use 1.2. □

1.5 Proposition. *The set $\mathbb{R}$ of all real numbers is uncountable.*

Proof. The proof uses the **Cantor diagonal argument**. We use the property that every real number can be expressed in decimal forms, see 1.23.

Suppose that set of all real numbers in the interval $[0,1]$ is countable, and is enumerated as a sequence $\{a_i \mid i \in \mathbb{Z}^+\}$. Let us write

$$\begin{aligned}
a_1 &= 0.a_{1,1}a_{1,2}a_{1,3}\ldots \\
a_2 &= 0.a_{2,1}a_{2,2}a_{2,3}\ldots \\
a_3 &= 0.a_{3,1}a_{3,2}a_{3,3}\ldots \\
&\ \vdots
\end{aligned}$$

Choose a number $b = 0.b_1b_2b_3\ldots$ such that $b_n \neq 0$, $b_n \neq 9$, and $b_n \neq a_{n,n}$, namely, pick $b_n = \min\{m \in \mathbb{Z} \mid 0 < m < 9, m \neq a_{n,n}\}$. The reason for choosing

$b_n \neq 0$ and $b_n \neq 9$ is to avoid real numbers whose decimal presentations are not unique, such as $\frac{1}{2} = 0.5000\ldots = 0.4999\ldots$, see 1.23. Then $b \neq a_n$ for all n. Thus the number b is not in the above table, a contradiction. $\square$

Example. Two intervals $[a, b]$ and $[c, d]$ on the real number line are equivalent. The bijection can be given by a map such as $x \mapsto y = \frac{d-c}{b-a}(x - a) + c$. Similarly, any two intervals (a, b) and (c, d) are equivalent. See Fig. 1.6.

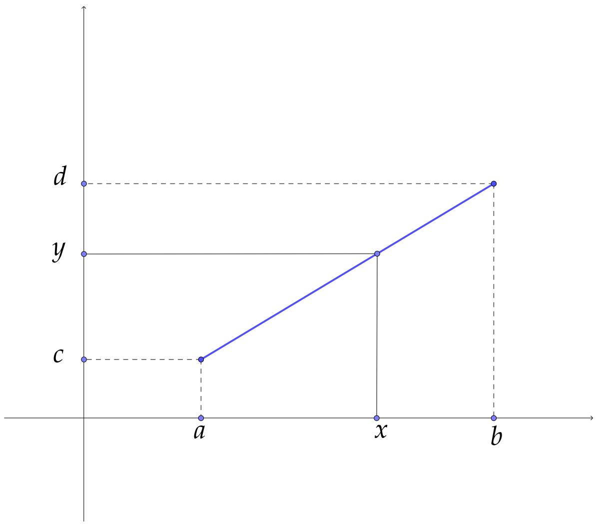


Figure 1.6: Any two intervals $[a, b]$ and $[c, d]$ are set-equivalent.

1.7 Example. The interval $(-1, 1)$ is equivalent to $\mathbb{R}$ via a map related to the tan function:

$$x \mapsto y = \frac{x}{\sqrt{1 - x^2}}.$$

We easily calculate the inverse map

$$\mathbb{R} \mapsto (-1, 1)$$
$$y \mapsto x = \frac{y}{\sqrt{1 + y^2}}.$$

See Fig. 1.8.

There is a simple injective map from a set S to the collection 2^S of all subsets of S, namely $a \mapsto \{a\}$. However there is no surjective map:

1.9 Theorem. *For any set $S \neq \emptyset$ there is no surjective map from S to 2^S. Thus a non-empty set cannot contain the collection of all of its subsets. In particular, there does not exist the set of all sets.*

Proof. Let ϕ be any map from S to 2^S. Let $X = \{a \in S \mid a \notin \phi(a)\}$. Suppose that there is $x \in S$ such that $\phi(x) = X$. If $x \in X$ then $x \notin \phi(x) = X$, a contradiction.

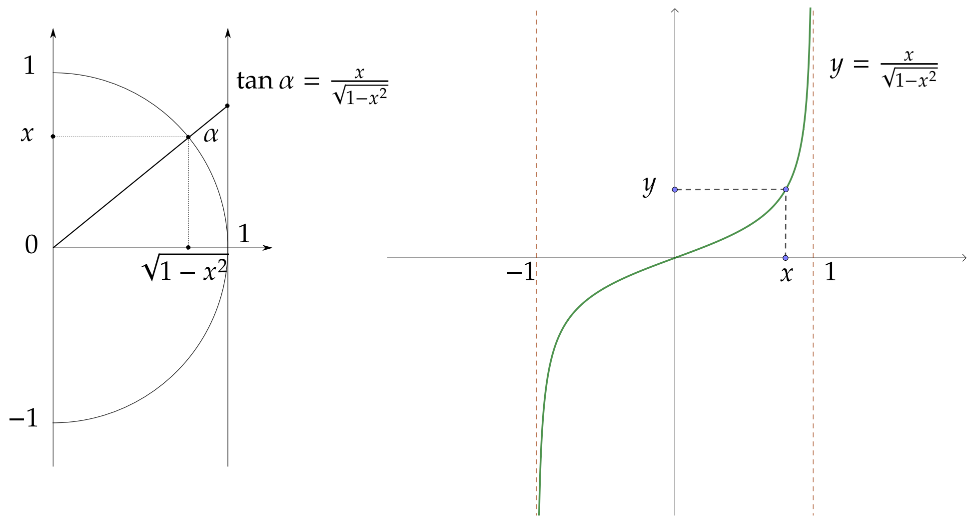


Figure 1.8: The interval $(-1, 1)$ is set-equivalent to $\mathbb{R}$.

If $x \notin X$ then $x \in \phi(x) = X$, also a contradiction. Therefore there is no $x \in S$ such that $\phi(x) = X$, hence ϕ is not surjective.

The above result implies that S cannot contain 2^S, for otherwise there are surjective maps from S onto 2^S such as

$$S \longrightarrow 2^S$$
$$a \longmapsto \begin{cases} a, & a \in 2^S \\ S, & a \notin 2^S. \end{cases}$$

□

The Axiom of choice

Proposition (Axiom of choice). *Given a collection of non-empty sets, there is a function defined on this collection, called a* **choice function**, *associating each set in the collection with an element of that set.*

Intuitively, a choice function "chooses" an element from each set in a given collection of non-empty sets. The Axiom of choice allows us to make infinitely many arbitrary choices.

Example. In the proof of 1.4, we choose enumerations of infinitely many countable sets, so we are employing the Axiom of choice.

In the proof of 1.3 we are able to use the well-ordered property of $\mathbb{Z}^+$ to specify a choice from each set, instead of making an arbitrary choice from each set, thus the use of the Axiom of choice is avoided. [1]

[1] Bertrand Russell explained that choosing one shoe from each pair of shoes in an infinite collection of pairs of shoes does not need the Axiom of choice, because in a pair of shoes the left shoe is different from the right shoe so we can define a choice such as for all pairs of shoes

In this course it is not required to specify whether the Axiom of choice is being used, or whether its use could be avoided.

The Axiom of choice is often used in constructions of functions, sequences, nets, see an example at 5.3. A common application is the use of the product of an infinite family of sets – the Cartesian product, discussed below. The Axiom of choice is needed for many important results in mathematics, such as the Tikhonov theorem in Topology, the Hahn–Banach theorem and Banach–Alaoglu theorem in Functional Analysis, the existence of a Lebesgue un-measurable set in Real Analysis, ….

Zorn lemma is often a convenient form of the Axiom of choice.

Proposition (Zorn lemma). *If any totally ordered subset of an ordered set X has an upper bound then X has a maximal element.*

Example (every vector space has a vector basis). In a vector space, a set of vectors is linearly independent if one vector could not be written as a linear combination, with coefficients not all zeros, of finitely many other vectors in the set. A vector basis for the vector space is a maximal set (under set inclusion) of linearly independent vectors.

If a vector basis exists, then any vector can be written as a linear combination of finitely many vectors belonging to the basis, by maximality.

Given a vector space, consider the collection F of all linearly independent sets of vectors with the order of set inclusion. Suppose that G is a totally ordered subset of F. Let $A = \bigcup_{B\in G} B$. It can be checked that $A \in F$ and A is an upper bound of G. By Zorn lemma F has a maximal element.

For more on this topic the reader may consult [End77, p. 151], [HJ99, p. 137].

Cartesian product

A map $f : I \to S$ where S is a collection of sets is called an **indexed collection** or a **family** [1] of sets. We often write $f_i = f(i)$, and denote the indexed collection f by $(f_i)_{i\in I}$ or $\{f_i\}_{i\in I}$. Notice that it can happen that $f_i = f_j$ for some $i \neq j$.

Example. A sequence of elements in a set A is a collection of elements of A indexed by the set $\mathbb{Z}^+$ of positive integer numbers, written as $(a_n)_{n\in\mathbb{Z}^+}$.

Let $(A_i)_{i\in I}$ be a collection of non-empty sets indexed by a set I. The **Cartesian product** [2] $\prod_{i\in I} A_i$ of this indexed collection is defined to be the collection of all maps $a : I \to \bigcup_{i\in I} A_i$ such that $a(i) \in A_i$ for every $i \in I$. In general, the

choose the left shoe. But usually in a pair of socks the two socks are identical, so choosing one sock from each pair of socks is arbitrary, so we need the Axiom of choice to be able to choose one sock from each pair of socks in an infinite collection of pairs of socks.

[1] họ

[2] tích Descartes

existence of such a map is a consequence of the Axiom of choice. An element a of $\prod_{i\in I} A_i$ is often denoted by $(a_i)_{i\in I}$, with $a_i = a(i) \in A_i$ being the coordinate of index i, in analogy with the finite product case.

Example. The set $\prod_{i\in\mathbb{Z}^+} \mathbb{R}$, also denoted by $\mathbb{R}^{\mathbb{Z}^+}$, is the set of all sequences of real numbers. Any element $a \in \mathbb{R}^{\mathbb{Z}^+}$ can be written as $a = (a_1, a_2, \ldots, a_n, \ldots)$ with $a_n \in \mathbb{R}$.

* More on sets

We discuss briefly several further items.

The notion of "ordered pair" used previously can be defined in terms of sets. Given a set S, an **ordered pair** of two elements a and b of S, written (a, b), can be defined as the set $\{a, \{a, b\}\}$. Thus $(b, a) = \{b, \{b, a\}\} \neq (a, b)$.

Then the **Cartesian product** a set A with a set B is the set of all ordered pairs (a, b) such that a is in A and b is in B (so both are in $A \cup B$), written as $A \times B = \{(a, b) \mid a \in A, b \in B\}$.

The notion of "map" can also be defined in terms of sets. A **map** or a **function** f from a set A to a set B is a subset of the set $A \times B$ such that for each $a \in A$ there is a unique $b \in B$ such that $(a, b) \in f$. We usually write $b = f(a)$, and thus a map $f : A \to B$ is determined by its graph $\{(a, f(a)) \mid a \in A\} \subset A \times B$.

For more on naive set theory the reader may consult [Hal74].

There are issues in naive set theory.

Example (Russell paradox). Consider the set $S = \{x \mid x \notin x\}$, the set of all sets which are not members of themselves. Whether $S \in S$ or not is undecidable, since answering either yes or no to this question leads to contradiction.

Russell paradox indicates that using the notion of set without definition may lead to contradiction. [1]

Efforts have been made since the beginning of the 20th century to resolve these issues.

Currently the Zermelo–Fraenkel–Choice axiom (ZFC) system of axioms is commonly used. In this system, there is the set of all subsets of a set, there is no set of all sets, no set can be a member of itself, thus Russell paradox is avoided. The existence of the set of natural numbers is included in this system. From the set of natural numbers the sets of integers, the set of rationals, and the set of real numbers are constructed. For more, see [End77], [HJ99].

In another system, the Von Neumann–Bernays–Godel system of axioms, a more general notion than set, called "class", is used. In this system, the class of all subsets of a set is a set, the class of all sets is not a set, no set can be a

[1] Discovered in 1901 by Bertrand Russell. A popular version of this paradox is the barber paradox: In a village there is a barber; the job of the barber is to do hair cut for a villager if and only if the villager does not cut his hair himself. Consider the set of all villagers who had hair cut by the barber. Is the barber a member of that set?

member of itself. This system used to be commonly adopted in textbooks in General Topology. Using the terms class or collection circumvents the issue of existences of sets. For more, see [Dug66, p. 18–20].

With this foundation, the rest of our study consists of no further axiom or undefined notion.

Problems

1.10. Let f be a map. Check that:

(a) $f(\bigcup_i A_i) = \bigcup_i f(A_i)$.

(b) $f(\bigcap_i A_i) \subset \bigcap_i f(A_i)$. If f is injective (one-one) then equality happens.

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

(d) $f^{-1}(\bigcap_i A_i) = \bigcap_i f^{-1}(A_i)$.

1.11. Let f be a map. Check that:

(a) $f(f^{-1}(A)) \subset A$. If f is surjective (onto) then equality happens.

(b) $f^{-1}(f(A)) \supset A$. If f is injective then equality happens.

1.12. Prove that any infinite set contains a countably infinite subset.

1.13. Show that if A is infinite and B is countable then $A \cup B$ is equivalent to A.

1.14. Explain why these sets of real numbers are set-equivalent: $(0, 1)$, $[0, 1)$, $[1, 1]$, $(-\infty, 1)$, $(-\infty, 1]$, $(-\infty, \infty)$.

1.15. Give another proof of 1.4 by checking that the map $\mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+$, $(m, n) \mapsto 2^m 3^n$ is injective.

1.16. ✓ Check that $\mathbb{Q}^n$ is countable.

1.17. Show that if A has n elements then $\left|2^A\right| = 2^n$.

1.18. Show that the set of all functions $f : A \to \{0, 1\}$ is equivalent to 2^A.

1.19. A real number α is called an algebraic number if it is a root of a polynomial with integer coefficients. Show that the set of all algebraic numbers is countable.

A real number which is not algebraic is called transcendental. For example it is known that π and e are transcendental. Show that the set of all transcendental numbers is uncountable.

1.20. A continuum set is a set which is set-equivalent to $\mathbb{R}$. Show that a countable union of continuum sets is a continuum set.

1.21 (Cantor set). Deleting the open interval $(\frac{1}{3}, \frac{2}{3})$ from the interval of real numbers $[0, 1]$, one gets a space consisting of two intervals $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Continuing, delete the intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$. In general on each of the remaining intervals, delete the middle open interval of $\frac{1}{3}$ the length of that interval. The Cantor set is the set of remaining points. It can be described as the set of real numbers $\sum_{n=1}^{\infty} \frac{a_n}{3^n}$, $a_n = 0, 2$. In other words, it is the set of real numbers in $[0, 1]$ which in base 3 could be written without the digit 1.

Show that the total length of the deleted intervals is 1. Is the Cantor set countable?

1.22 (Cantor–Schroeder–Bernstein theorem). * We prove: If A is set-equivalent to a subset of B and B is set-equivalent to a subset of A then A and B are set-equivalent.

Suppose that $f : A \mapsto B$ and $g : B \mapsto A$ are injective maps. Let $A_0 = A$ and $B_0 = B$. For $n \in \mathbb{Z}$, $n \geq 0$, let $B_{n+1} = f(A_n)$ and $A_{n+1} = g(B_n)$.

(a) Show that $A_{n+1} \subset A_n$ and $B_{n+1} \subset B_n$.

(b) Show that A_{n+2} is set-equivalent to A_n, and $A_n \setminus A_{n+1}$ is set-equivalent to $A_{n+2} \setminus A_{n+3}$.

(c) By writing

$$A = [(A_0 \setminus A_1) \cup (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup (A_3 \setminus A_4) \cup \cdots] \cup \left(\bigcap_{n=1}^{\infty} A_n\right)$$

$$A_1 = [(A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup (A_3 \setminus A_4) \cup (A_4 \setminus A_5) \cup \cdots] \cup \left(\bigcap_{n=1}^{\infty} A_n\right)$$

deduce that A is set-equivalent to A_1.

For more, see [KF75, p. 17], [End77, p. 148].

1.23. Give a proof that any real number could be written in base d with any $d \in \mathbb{Z}$, $d \geq 2$. More specifically, any positive real number x can be written as

$$x = a_n \ldots a_1 a_0 . b_1 b_2 b_3 \cdots = a_n d^n + \cdots + a_1 d^1 + a_0 + \frac{b_1}{d} + \frac{b_2}{d^2} + \frac{b_3}{d^3} + \cdots$$

where $a_i \in \mathbb{Z}$, $b_i \in \mathbb{Z}$, $0 \leq a_i \leq d-1$, $0 \leq b_i \leq d-1$.

However two forms in base d can represent the same real number, as seen in 1.5. This happens only if starting from certain digits after the dot, all digits of one form are 0 and all digits of the other form are $d-1$. For $d = 2$ this is the binary form, and for $d = 10$ this is the decimal form of real numbers. (This result is used in 1.5.)

1.24 ($\mathbb{R}^n$ is set-equivalent to $\mathbb{R}$). Here we prove that $\mathbb{R}^2$ is equivalent to $\mathbb{R}$, in other words, a plane is set-equivalent to a line, or the set $\mathbb{C}$ of complex numbers is set-equivalent to the set $\mathbb{R}$ of real numbers. As a corollary, $\mathbb{R}^n$ is set-equivalent to $\mathbb{R}$.

Let us construct a map from $[0,1) \times [0,1)$ to $[0,1)$ as follows: the pair of two real numbers in decimal forms $0.a_1a_2\ldots$ and $0.b_1b_2\ldots$ corresponds to the real number $0.a_1b_1a_2b_2\ldots$. In view of 1.23, we only allow decimal presentations in which not all digits are 9 starting from a certain digit. Check that this map is injective.

1.25. We prove that the set $2^{\mathbb{Z}^+}$ of all subsets of $\mathbb{Z}^+$ is set-equivalent to $\mathbb{R}$.

(a) Show that $2^{\mathbb{Z}^+}$ is set-equivalent to the set of all sequences of binary digits. Deduce that there is an injective map from $[0,1]$ to $2^{\mathbb{Z}^+}$.

(b) Consider a map $f : 2^{\mathbb{Z}^+} \to [0,2]$, for each binary sequence $a = a_1a_2a_3\ldots$, if starting from a certain digit all digits are 1 then let $f(a) = 1.a_1a_2a_3\ldots$ in binary form, otherwise let $f(a) = 0.a_1a_2a_3\ldots$. Show that f is injective.

The statement that there does not exist a set that is strictly more than $\mathbb{Z}^+$ but is strictly less than $\mathbb{R}$ (meaning a set S for which there is an injective map from $\mathbb{Z}^+$ to S but there is not a bijective map, and there is an injective map from S to $\mathbb{R}$ but there is not a bijective map) is called the **Continuum hypothesis** and is known to be independent from the ZFC axioms [HJ99, p. 268].

1.26 (transfinite induction principle). An ordered set S is **well-ordered** [1] if every non-empty subset A of S has a smallest element, i.e. $\exists a \in A, \forall b \in A, a \leq b$. For examples, with the usual order $\mathbb{N}$ is well-ordered while $\mathbb{R}$ is not. Notice that a well-ordered set must be totally ordered. It is known that, based on the Axiom of choice, any set can be well-ordered.

Prove the following generalization of the principle of induction. Let A be a well-ordered set. Let $P(a)$ be a statement whose truth depends on $a \in A$. Suppose that if $P(a)$ is true for all $a < b$ then $P(b)$ is true. Then $P(a)$ is true for all $a \in A$.

For more, see for examples [End77, p. 174], [HJ99, p. 114], [Dug66, p. 40].

1.27 (cardinality). We briefly introduce a notion for comparing "sizes" of sets. If there is a bijective map from a set A onto a set B, we say that the **cardinality** [2] of A is equal to the cardinality of B. Thus two sets have same cardinality if and only if they are set-equivalent. If there is an injective map from a set A to a set B, we say that the cardinality of A is less than or equal to the cardinality of B. Check the followings.

(a) Having same cardinality has the properties of an equivalence relation.

(b) Consider equivalence classes of sets, having less than or equal cardinality has the properties of an order relation.

(c) Among infinite sets, the cardinality of the set $\mathbb{Z}^+$ is the smallest. There is no largest cardinality.

For more, see for examples [End77, p. 128], [HJ99, p. 65].

[1] được sắp tốt
[2] lực lượng

2 Topological space

Topology is a general setting for discussions involving convergence and continuity.

Example. In Calculus, a sequence of real numbers $(x_n)_{n\in\mathbb{Z}^+}$ is convergent to a real number x if the distance between x_n and x is arbitrarily small for n sufficiently large.

It is equivalent to this statement: x_n belongs to arbitrary open interval containing x for n sufficiently large.

The second formulation does not explicitly employ distance. Thus it is possible to discuss convergence without distance, it is sufficient to use the notion of open intervals.

Metric space

More generally let us recall how continuity arises in metric spaces.

Recall that, briefly, a metric space is a set equipped with a distance between every two points. Namely, a metric space is a set X with a map $d : X \times X \mapsto \mathbb{R}$ such that for all $x, y, z \in X$:

(a) $d(x, y) \geq 0$ (distance is non-negative),

(b) $d(x, y) = 0 \iff x = y$ (distance is zero if and only if the two points coincide),

(c) $d(x, y) = d(y, x)$ (distance is symmetric),

(d) $d(x, y) + d(y, z) \geq d(x, z)$ (triangular inequality).

Example (normed space). Recall that a normed space [1] is briefly a vector space equipped with lengths of vectors. Namely, a normed space over the field of real numbers is a set X with a structure of a vector space over the real numbers and a real function $X \to \mathbb{R}$, $x \mapsto \|x\|$, called a **norm** [2], satisfying:

(a) $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$ (length is non-negative),

(b) $\|cx\| = |c| \|x\|$ for $c \in \mathbb{R}$ (length is proportionate to vector),

(c) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

A normed space is canonically a metric space with metric $d(x, y) = \|x - y\|$.

[1] không gian định chuẩn
[2] chuẩn

In a metric space (X, d), a ball is a set of the form $B(x, r) = \{y \in X \mid d(y, x) < r\}$ where $r \in \mathbb{R}$, $r > 0$.

A subset U of X is open if for all x in U there is $\epsilon > 0$ such that $B(x, \epsilon)$ is contained in U. This is equivalent to saying that ***a non-empty open set is a union of balls***.

A sequence $(x_n)_{n \in \mathbb{Z}^+}$ is convergent to an element x in X if for any ball at x, for n sufficiently large, x_n is in that ball:

$$\forall \epsilon > 0, \exists n_0 \in \mathbb{Z}^+, \forall n \geq n_0, x_n \in B(x, \epsilon).$$

This is equivalent to the statement that for any open set containing x, for n sufficiently large, x_n is in that open set:

$$\forall U \in \{U \subset X \mid U \ni x, U \text{ open}\}, \exists n_0 \in \mathbb{Z}^+, \forall n \geq n_0, x_n \in U.$$

In this form convergence only uses open sets.

A map $f : (X, d_X) \to (Y, d_Y)$ is continuous at $x \in X$ if and only if the distance between $f(y)$ and $f(x)$ is arbitrarily small if the distance between y and x is sufficiently small:

$$\forall \epsilon > 0, \exists \delta > 0, \forall y \in X, d_X(y, x) < \delta \implies d_Y(f(y), f(x)) < \epsilon,$$

equivalently,

$$\forall \epsilon > 0, \exists \delta > 0, \forall y \in B_X(x, \delta), f(y) \in B_Y(f(x), \epsilon),$$

or even shorter,

$$\forall \epsilon > 0, \exists \delta > 0, f(B_X(x, \delta)) \subset B_Y(f(x), \epsilon).$$

This is equivalent to

$$\forall U \in \{U \subset Y \mid U \ni f(x), U \text{ open}\}, \exists V \in \{V \subset X \mid V \ni x, V \text{ open}\}, f(V) \subset U.$$

In this form continuity only uses open sets.

Example (different metrics with same convergence and same continuity). Common norms in $\mathbb{R}^n$ are

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, p \in \mathbb{R}, p \geq 1,$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

In Analysis it is known that with these norms convergence of sequences are same and continuity of maps are same, see 2.25 and 2.26. An explanation is that although these norms have different balls, these balls generate the same collection of open sets, see Fig. 2.1, Fig. 2.3, and 2.2. Thus if we are only interested in certain properties such as convergence and continuity then which norm is being used does not matter.

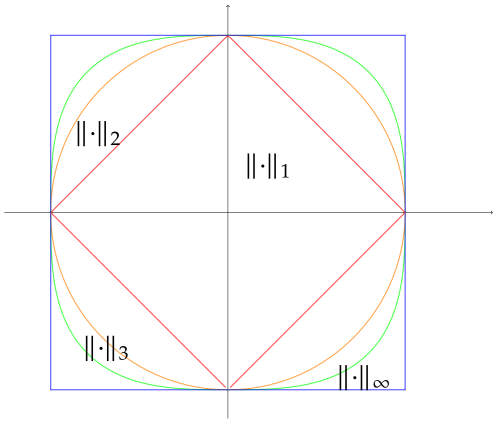


Figure 2.1: Unit balls with respect to different norms on $\mathbb{R}^2$.

Example. On the two dimensional unit sphere S^2 in $\mathbb{R}^3$, such as the surface of the Earth, we certainly use the notion of convergence and continuity, but it is open to debate whether it is appropriate to use the Euclidean metric of $\mathbb{R}^3$ for S^2, and if not, which metric should be used.

In summary, we see that in metric spaces convergence and continuity can be expressed through open sets, and there are situations where which metric is being used is not relevant or not clear.

Open sets in metric spaces have several simple properties. ***A union of open sets is an open set***. The intersection of two open sets is an open set. Indeed, let A and B be open sets. Let $x \in A \cap B$. Since $x \in A$ there is a ball $B(x, r) \subset A$, and since $x \in B$ there is a ball $B(x, s) \subset B$. Let $t = \min\{r, s\}$, then $B(x, t) \subset A$ and $B(x, t) \subset B$, so $B(x, t) \subset A \cap B$. Thus $A \cap B$ is open. By induction, ***the intersection of any finite collection of open sets is an open set***. However, ***an infinite intersection of open sets can be not open***, for example $\bigcap_{n=1}^{\infty} B(x, \frac{1}{n}) = \{x\}$.

Topology

When we discuss dependence of an object on another object we are often interested in continuity of the dependence. This notion is often understood as that the sets of variations of the dependent object can be controlled by the sets of variations of the independent object. The sets that appear in this discussion are called open sets. Briefly, ***a topology is a system of open sets***.

In metric space the notion of open sets is expressed using distance. To generalize this notion without the present of distance, we directly specify which sets are open sets. We shall choose certain properties of open sets in metric spaces to retain, specifically: ***unions of open sets are open, finite intersections of open sets are open***. Although there can be other approaches for other purposes, since the early 20th century the following setting has shown to be sufficiently general and effective for many areas of mathematics:

Definition. *A* **topology** [1] *on a set X is a collection τ of subsets of X satisfying the following conditions:*

(a) The sets $\emptyset$ and X are elements of τ. In symbols: $\emptyset \in \tau, X \in \tau$.

(b) A union of elements of τ is an element of τ. In symbols:

$$\forall F \subset \tau, \bigcup_{O \in F} O \in \tau.$$

(c) A finite intersection of elements of τ is an element of τ. In symbols:

$$\forall F \in \{F \subset \tau \mid |F| < \infty\}, \bigcap_{O \in F} O \in \tau.$$

The elements of τ are called the **open sets** *of X in the topology τ.*

In short, ***a topology on a set X is a collection of subsets of X which includes $\emptyset$ and X and is "closed" under unions and finite intersections***.

A set X together with a topology τ is called a **topological space**, denoted by(X, τ) or X alone if we do not need to specify the topology. An element of X is often called a **point**.

Example. On any set X there is the **trivial topology** [2] $\{\emptyset, X\}$. There is also the **discrete topology** [3] whereas any subset of X is open. Thus on a set there can be many topologies.

Example. Let $X = \{1, 2, 3\}$. The collection $\tau = \{\emptyset, \{1, 2\}, \{2, 3\}, \{2\}, \{1, 2, 3\}\}$ is a topology on X.

Notice that the statement "intersection of any finitely many open sets is open" is equivalent to the statement "intersection of any two open sets is open", by induction.

Example (topology generated by metric). A metric space is canonically a topological space with the topology generated by the metric, i.e. non-empty open sets are unions of balls. ***When we speak about topology on a metric space we mean the topology generated by the metric***.

[1] tôpô
[2] tôpô hiển nhiên
[3] tôpô rời rạc

Example (topology generated by norm). A normed space is canonically a metric space with metric $d(x, y) = \|x - y\|$. Therefore a normed space is canonically a topological space with the topology generated by the metric from the norm.

Example (Euclidean topology). In $\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_i \in \mathbb{R}\}$, the Euclidean norm of a point $x = (x_1, x_2, \dots, x_n)$ is $\|x\| = \left[\sum_{i=1}^n x_i^2\right]^{1/2}$. The topology generated by this norm is called the **Euclidean topology** [1] of $\mathbb{R}^n$.

Special subsets of topological spaces

The following notions generalize the corresponding notions in metric spaces.

A subset of a topological space is called a **closed set** if its complement is an open set. In symbols, the set $A \subset (X, \tau)$ is a closed set in the topogical space (X, τ) if $X \setminus A \in \tau$.

Proposition. *In a topological space X:*

(a) $\emptyset$ and X are closed.

(b) A finite union of closed sets is closed.

(c) An intersection of closed sets is closed.

Proof. These properties are deduced from the set identities

$$\bigcup_U (X \setminus U) = X \setminus \left(\bigcap_U U\right),$$

and

$$\bigcap_U (X \setminus U) = X \setminus \left(\bigcup_U U\right).$$

□

Remark. Being closed and being open are not exclusive, in possible contrast to the uses of the terms outside of topology. A subset of space, such as $\emptyset$ and the whole space itself, can be both closed and open in that space, or can be neither closed nor open. Our experience in Euclidean spaces, where except the empty set and the whole space, being open and being closed are indeed exclusive, see 4.5, is not correct in general.

Let X be a topological space and let A be a subset of X. A point x in X is said to be:

[1] tôpô Euclid

(a) An **interior point** [1] of A in X if there is an open set of X containing x that is contained in A. In symbols:

$$\exists U \in \{U \in \tau \mid x \in U\}, U \subset A.$$

(b) A **contact point** [2] (or point of closure) of A in X if any open set of X containing x contains a point of A. In symbols,

$$\forall U \in \{U \in \tau \mid x \in U\}, U \cap A \neq \emptyset.$$

(c) A **limit point** [3] (or cluster point, or accumulation point) of A in X if any open set of X containing x contains a point of A other than x. In symbols,

$$\forall U \in \{U \in \tau \mid x \in U\}, U \cap (A \setminus \{x\}) \neq \emptyset.$$

(d) A **boundary point** [4] of A in X if any open set of X containing x contains a point of A and a point of the complement of A. In other words, a boundary point of A is a contact point of both A and the complement of A. In symbols,

$$\forall U \in \{U \in \tau \mid x \in U\}, U \cap A \neq \emptyset, U \cap (X \setminus A) \neq \emptyset.$$

(e) The **interior** [5] of A in X, denoted by $\mathring{A}$ or $\mathrm{int}(A)$, is the set of all interior points of A.

(f) The **closure** [6] of A in X, denoted by $\overline{A}$ or $\mathrm{cl}(A)$, is the set of all contact points of A.

(g) The **boundary** [7] of A in X, denoted by ∂A, is the set of all boundary points of A in X.

(h) A **neighborhood** [8] of a point $x \in X$ is a subset of X which contains an open set containing x. Note that a neighborhood does not need to be open. [9]

Example. On the Euclidean line $\mathbb{R}$, consider the subset $A = [0, 1) \cup \{2\}$. Its interior is $\mathrm{int} A = (0, 1)$, the closure is $\mathrm{cl} A = [0, 1] \cup \{2\}$, the boundary is $\partial A = \{0, 1, 2\}$, the set of all limit points is $[0, 1]$.

[1] điểm trong
[2] điểm dính
[3] điểm tụ
[4] điểm biên
[5] phần trong
[6] bao đóng
[7] biên
[8] lân cận
[9] Not everyone uses this convention, for examples [Kel55] uses this convention but [Mun00] requires a neighborhood to be open.

Bases of a topology

Definition. *Given a topology, a collection of open sets is a* **basis** [1] *for that topology if every non-empty open set is a union of members of that collection.*

More concisely, let τ be a topology of X, then a collection $B \subset \tau$ is called a basis for τ if for any $\emptyset \neq V \in \tau$ there is $C \subset B$ such that $V = \bigcup_{O \in C} O$.

So a basis of a topology is a subset of the topology that generates the entire topology via unions.

Example. On $X = \{1,2,3\}$ the topology $\tau = \{\emptyset, \{1,2\}, \{2,3\}, \{2\}, \{1,2,3\}\}$ has a basis $\{\{1,2\}, \{2,3\}, \{2\}\}$. Specifying this basis is enough to determine the topology.

Example. The trivial topology on X has a basis consisting of only the set X, that is, $\{X\}$.

Example. The discrete topology on X has a basis consisting of sets containing only one element of X, that is, $\{\{x\} \mid x \in X\}$.

Example. In a metric space the collection of all balls is a basis for the topology generated by the metric.

2.2 Example. The Euclidean plane has a basis consisting of all open disks. Observe that any disk contains a square with the same center, see Fig. 2.3. Thus any open set is also a union of open squares, hence the Euclidean plane also has a basis consisting of all open squares.

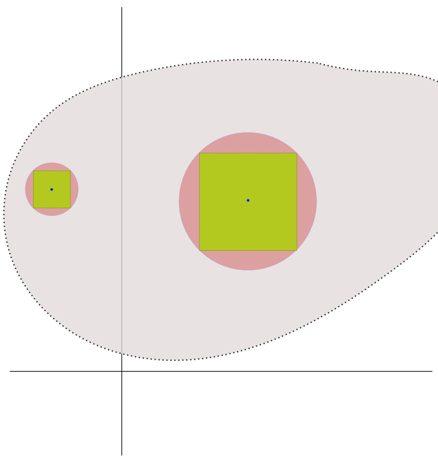


Figure 2.3: The set of squares is a basis for the Euclidean topology of the plane.

In a similar manner other shapes can be used as bases for the Euclidean topology, see 2.22, 2.26, and Fig. 2.1.

Definition. *A collection $S \subset \tau$ is called a* **subbasis** [2] *for the topology τ if the collection of all finite intersections of members of S is a basis for τ.*

[1] cơ sở
[2] tiền cơ sở

Clearly a basis for a topology is also a subbasis for that topology. Briefly, given a topology, a subbasis is a subset of the topology that can generate the entire topology by finite intersections and then by unions.

Example. On $X = \{1,2,3\}$ the topology $\tau = \{\emptyset, \{1,2\}, \{2,3\}, \{2\}, \{1,2,3\}\}$ has a subbasis $\{\{1,2\}, \{2,3\}\}$.

2.4 Example. The collection of all open rays, that are, sets of the forms (a, ∞) and $(-\infty, a)$, is a subbasis for the Euclidean topology of $\mathbb{R}$.

Generating topologies

Suppose we want certain subsets of a set to be open sets, can we find such a topology? Of course the discrete topology for which every subset is open satisfies the requirement, so the goal is to find a minimal topology for our purpose.

On a set X we can compare topologies by using the order relation by set inclusion. Namely, let τ_1 and τ_2 be two topologies on X, if $\tau_1 \subset \tau_2$ we say that τ_2 is **finer** [1] (or stronger, bigger) than τ_1, and τ_1 is **coarser** [2] (or weaker, smaller) than τ_2.

Example. On any set the trivial topology is the coarsest topology and the discrete topology is the finest one.

2.5 Proposition. *If S be a collection of subsets of X, then the intersection of all topologies of X containing S is the smallest topology that contains S, called* **the topology generated by the collection of subsets** *S.*

Proof. We can check easily that an intersection of topologies on X is a topology on X, Problem 2.19. □

This is remarkable: ***any collection of subsets generates a topology***. If we want certain sets to be open, there is a minimal topology for that.

Moreover this topology can be built as follows.

2.6 Proposition. *Let S be a collection of subsets of X. The collection τ consisting of $\emptyset$, X, and all unions of finite intersections of members of S is the smallest topology on X containing S – the topology generated by S.*

Concisely, letting $B = \{\bigcap_{O \in I} O \mid I \subset S, |I| < \infty\}$, then the topology generated by S is $\tau = \{\emptyset, X, \bigcup_{U \in F} U \mid F \subset B\}$. The collection S is a subbasis for τ and the collection B is the corresponding basis for τ.

[1] mịn hơn
[2] thô hơn

Proof. Any topology of X containing S must contain τ. We only need to check that τ is a topology. [1]

We check that τ is closed under unions. Let $\sigma \subset \tau$, consider $\bigcup_{A\in\sigma} A$. We write $\bigcup_{A\in\sigma} A = \bigcup_{A\in\sigma} \left(\bigcup_{U\in F_A} U\right)$, where $F_A \subset B$. Since

$$\bigcup_{A\in\sigma}\left(\bigcup_{U\in F_A} U\right) = \bigcup_{U\in(\bigcup_{A\in\sigma} F_A)} U,$$

and since $\bigcup_{A\in\sigma} F_A \subset B$, we conclude that $\bigcup_{A\in\sigma} A \in \tau$.

We check that τ is closed under intersections of two elements. Let $\bigcup_{U\in F} U$ and $\bigcup_{V\in G} V$ be two elements of τ, where $F, G \subset B$. We can write

$$\left(\bigcup_{U\in F} U\right) \cap \left(\bigcup_{V\in G} V\right) = \bigcup_{U\in F, V\in G} (U\cap V).$$

Let $J = \{U \cap V \mid U \in F, V \in G\}$. Then $J \subset B$, and we can write

$$\left(\bigcup_{U\in F} U\right) \cap \left(\bigcup_{V\in G} V\right) = \bigcup_{W\in J} W,$$

showing that $(\bigcup_{U\in F} U) \cap (\bigcup_{V\in G} V) \in \tau$. □

Example. Let $X = \{1,2,3,4\}$. The set $\{\{1\},\{2,3\},\{3,4\}\}$ generates the topology

$$\{\emptyset, \{1\}, \{3\}, \{1,3\}, \{2,3\}, \{3,4\}, \{1,2,3\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\}.$$

A basis for this topology is $\{\{1\},\{3\},\{2,3\},\{3,4\}\}$.

Example. Let $(X, \leq)$ be a totally ordered set. The collection of subsets of the forms $\{\beta \in X \mid \beta < \alpha\}$ and $\{\beta \in X \mid \beta > \alpha\}$ generates a topology on X, called the **ordering topology**.

Example. The Euclidean topology on $\mathbb{R}$ is the ordering topology with respect to the usual order of real numbers. Compare Example 2.4.

Problems

2.7. ✓ The **finite complement topology** on X consists of the empty set and all subsets of X whose complements are finite. Check that this is indeed a topology.

2.8. ✓ The **countable complement topology** on X consists of the empty set and all subsets of X whose complements are countable. Check that this is indeed a topology. Compare the countable complement topology to the finite complement topology.

[1] In several textbooks the explicit inclusion of $\emptyset$ and X to τ is replaced by conventions on unions and intersections of empty collections.

2.9. Let X be a set and $p \in X$. Show that the collection consisting of $\emptyset$ and all subsets of X containing p is a topology on X. This topology is called the **particular point topology** on X, denoted by PPX_p. Describe the closed sets in this space.

2.10 (Sorgenfrey line). The collection of all intervals of the form $[a, b)$ generates a topology on $\mathbb{R}$. Compare this topology with the Euclidean topology.

2.11. ✓Show that:

(a) The interior of A in X is the largest open subset of X that is contained in A.

(b) A subset is open if and only if all of its points are interior points.

(c) The closure of A in X is the smallest closed subset of X containing A.

(d) A subset is closed if and only if it contains all of its contact points.

2.12. Let X be a topological space and $A \subset X$.

(a) Show that $\overline{A}$ is the disjoint union of $\mathring{A}$ and ∂A.

(b) Show that X is the disjoint union of $\mathring{A}$, ∂A, and $X \setminus \overline{A}$.

2.13. In a metric space X, a point $x \in X$ is a limit point of the subset A of X if and only if there is a sequence in $A \setminus \{x\}$ converging to x. (This is not true in general topological spaces, see 5.3.)

2.14. Find the closures, interiors and the boundaries of the interval $[0, 1)$ under the Euclidean, discrete and trivial topologies of $\mathbb{R}$.

2.15. Show that

(a) In a normed space, the boundary of the ball $B(x, r)$ is the sphere $\{y \mid \|x - y\| = r\}$, and so the ball $B'(x, r) = \{y \mid \|x - y\| \le r\}$ is the closure of $B(x, r)$.

(b) In a metric space, the boundary of the ball $B(x, r)$ is a subset of the sphere $\{y \mid d(x, y) = r\}$. Is the ball $B'(x, r) = \{y \mid d(x, y) \le r\}$ the closure of $B(x, r)$?

2.16. Let $O_n = \{k \in \mathbb{Z}^+ \mid k \ge n\}$. Check that $\{\emptyset\} \cup \{O_n \mid n \in \mathbb{Z}^+\}$ is a topology on $\mathbb{Z}^+$. Find the closure of the set $\{5\}$. Find the closure of the set of all even positive integers.

2.17. Show that any open set in the Euclidean line is a countable union of open intervals.

2.18. In the real number line with the Euclidean topology, is the Cantor set (see 1.21) closed or open, or neither? Find the boundary and the interior of the Cantor set.

2.19 (an intersection of topologies is a topology). ✓ Show that the intersection of a collection of topologies on a set X is a topology on X.

2.20. ✓ Let X be a topological space. A collection B of open sets is a basis if for each point x and each open set O containing x there is a U in B such that U contains x and U is contained in O.

2.21. Let B be a collection of subsets of a set X. Then $B \cup \{X\}$ is a basis for a topology on X if and only if the intersection of two members of B is either empty or is a union of some members of B. (In several textbooks to avoid adding the element X to B it is required that the union of all members of B is X.)

2.22. ✓ Check that:

(a) Two bases generate the same topology if and only if each member of one basis is a union of members of the other basis.

(b) Two bases generate the same topology if and only if whenever a point x belongs to an element U of one basis then there is an element V of the other basis such that $x \in V \subset U$.

2.23. In a metric space the set of all balls with rational radii is a basis for the topology. The set of all balls with radii $\frac{1}{2^m}$, $m \geq 1$ is another basis.

2.24 ($\mathbb{R}^n$ has a countable basis). ✓ The collection of all balls each with rational radius whose center has rational coordinates forms a basis for the Euclidean topology of $\mathbb{R}^n$.

2.25 (equivalent metrics). Let d_1 and d_2 be two metrics on X. If there are constant real numbers $\alpha > 0, \beta > 0$ such that $\alpha d_1 \leq d_2 \leq \beta d_1$ then in Analysis the two metrics are said to be equivalent. Show that two equivalent metrics generate same topologies.

2.26 (every norm in $\mathbb{R}^n$ generates the Euclidean topology). In $\mathbb{R}^n$ denote by $\|\cdot\|_2$ the Euclidean norm, and let $\|\cdot\|$ be any norm.

(a) Check that the map $x \mapsto \|x\|$ from $(\mathbb{R}^n, \|\cdot\|_2)$ to $(\mathbb{R}, \|\cdot\|_2)$ is continuous.

(b) Let S^{n-1} be the unit sphere under the Euclidean norm in $\mathbb{R}^n$. Show that the restriction of the map above to S^{n-1} has a maximum value β and a minimum value α. Hence $\alpha \leq \left\| \frac{x}{\|x\|_2} \right\| \leq \beta$ for all $x \neq 0$.

(c) Deduce that any two norms in $\mathbb{R}^n$ generate equivalent metrics, hence every norm in $\mathbb{R}^n$ generates the Euclidean topology.

The result that any two norms in a finite dimensional vector space are equivalent is often discussed in Analysis, see for example [TTV].

2.27. Let (X, d) be a metric space.

(a) Let $d_1(x, y) = \min\{d(x, y), 1\}$. Show that d_1 is a metric on X generating the same topology as that generated by d. Is d_1 equivalent to d?

(b) Let $d_2(x, y) = \frac{d(x,y)}{1+d(x,y)}$. Show that d_2 is a metric on X generating the same topology as the topology generated by d. Is d_2 equivalent to d? Is d_2 equivalent to d_1?

2.28. On $\mathbb{R}^2$, compare the Euclidean topology with the ordering topology with respect to the dictionary order.

2.29. On the set of all integer numbers $\mathbb{Z}$, consider arithmetic progressions

$$S_{a,b} = a + b\mathbb{Z},$$

where $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^+$.

(a) Show that these sets form a basis for a topology on $\mathbb{Z}$.

(b) Show that with this topology each set $S_{a,b}$ is closed.

3 Continuity

Continuous maps

As discussed in Section 2, in metric spaces, ***continuity of a map means the values of the map are inside any arbitrarily given open set provided the variables are inside a certain open set***. With a topology we specify a collection of sets to be the collection of open sets, therefore we can propose a generalization of continuity to topological spaces:

Definition. *Let X and Y be topological spaces. We say a map $f : X \to Y$ is* **continuous** *at a point x in X if for any open set U of Y containing $f(x)$ there is an open set V of X containing x such that $f(V)$ is contained in U. In symbols:*

$$\forall U \in \{U \in \tau \mid U \ni f(x)\}, \exists V \in \{V \in \tau \mid V \ni x\}, f(V) \subset U.$$

We say that a map is **continuous on the space** *if it is continuous at every point in the space.*

It is apparent that this definition is equivalent to the definition of continuity in metric spaces where the topologies are generated by the metric. In other words, if we look at a metric space as a topological space with the topology generated by the metric then continuity in the metric space is the same as continuity in the topological space. Therefore ***we inherit all results concerning continuity when the space is a metric space***.

Example. When $\mathbb{R}$ is equipped with the Euclidean topology then the sine function $\sin : \mathbb{R} \to \mathbb{R}$ is topologically continuous, since we know the sine function is continuous under the Euclidean metric, and the Euclidean metric generates the Euclidean topology.

Example. Let X and Y be topological spaces.

(a) The identity function, $\mathrm{id}_X : X \to X$, $x \mapsto x$, is continuous.

(b) The constant function, with a given $a \in Y$, $x \mapsto a$, is continuous.

(c) If Y has the trivial topology then any map $f : X \to Y$ is continuous.

(d) If X has the discrete topology then any map $f : X \to Y$ is continuous.

Theorem. *A map is continuous if and only if the inverse image of an open set is an open set.*

Proof. ($\Rightarrow$) Suppose that $f : X \to Y$ is continuous. Let U be an open set in Y. Let $x \in f^{-1}(U)$. Since f is continuous at x and U is an open neighborhood of

$f(x)$, there is an open set V_x containing x such that V_x is contained in $f^{-1}(U)$. Therefore $f^{-1}(U) = \bigcup_{x\in f^{-1}(U)} V_x$ is open.

($\Leftarrow$) Suppose that the inverse image of any open set is an open set. Let $x \in X$. Let U be an open neighborhood of $f(x)$. Then $V = f^{-1}(U)$ is an open set containing x, and $f(V)$ is contained in U. Therefore f is continuous at x. $\square$

It is often convenient to use the following property:

3.1 Proposition. *Suppose that $f : X \to Y$ and S is a subbasis for the topology of Y. Then f is continuous if and only if the inverse image of any element of S is an open set in X.*

This is Problem 3.11.

Example. A map from a topological space to a metric space is continuous if and only if the inverse image of any open ball is an open set.

A map $f : X \to \mathbb{R}$ where $\mathbb{R}$ has the Euclidean topology is continuous if and only if for any $a \in \mathbb{R}$ the sets $f^{-1}((-\infty, a))$ and the set $f^{-1}((a, \infty))$ are open.

Subspace

Let (X, τ) be a topological space and let Y be a subset of X. We want to define a topology on Y that can be naturally considered as being "inherited" from X. Thus any open set of X that is contained in Y should be considered open in Y. If an open set of X is not contained in Y then its restriction to Y should be considered open in Y. We can easily check that the collection of restrictions of the open sets of X to Y is a topology on Y.

Definition. *Let Y be a subset of the topological space X. The* **subspace topology** *on Y, also called the* **relative topology** [1] *with respect to X, is defined to be the collection of restrictions of the open sets of X to Y, that is, the collection $\{O \cap Y \mid O \in \tau\}$. With this topology we say that Y is a* **subspace** [2] *of X.*

In brief, a subset of a subspace Y of X is open in Y if and only if it is a restriction of an open set of X to Y.

Remark. An open or a closed subset of a subspace Y of a space X is not necessarily open or closed in X. For example, under the Euclidean topology of $\mathbb{R}$, the set $[0, \frac{1}{2})$ is open in the subspace $[0, 1]$, but is not open in $\mathbb{R}$. ***When we say that a set is open, we must know which topology we are using***.

3.2 Proposition. *Let X be a topological space and let $Y \subset X$. The subspace topology on Y is the coarsest topology on Y such that the inclusion map $i : Y \hookrightarrow X$, $x \mapsto x$ is*

[1] tôpô tương đối
[2] không gian con

continuous. We may say that the subspace topology on Y is the topology generated by the inclusion map from Y to X.

Proof. If O is a subset of X then $i^{-1}(O) = O \cap Y$. Thus the coarsest topology such that i is continuous is $\{O \cap Y \mid O \in \tau_X\}$, exactly the subspace topology of Y. □

Example (subspaces of a metric space). The notion of topological subspaces is compatible with the earlier notion of metric subspaces. Let (X, d) be a metric space and let $Y \subset X$. Then, as we know, Y is a metric space with the metric inherited from X, namely $d_Y = d_X|_{Y\times Y}$. A ball in Y is a set of the form, for $a \in Y$, $r > 0$:

$$B_Y(a,r) = \{y \in Y \mid d(y,a) < r\} = \{x \in X \mid d(x,a) < r\} \cap Y = B_X(a,r) \cap Y.$$

Thus a ball in Y is the restriction of a certain ball in X.

Any open set A in Y is the union of a collection of balls in Y:

$$A = \bigcup_{i\in I,\ a_i\in Y} B_Y(a_i, r_i) = \bigcup_{i\in I} (B_X(a_i, r_i) \cap Y) = \left(\bigcup_{i\in I} B_X(a_i, r_i)\right) \cap Y.$$

Thus an open set in Y is the restriction of a certain open set in X.

Conversely, if B is open in X, then for each $x \in B$ there is $r_x > 0$ such that $B_X(x, r_x) \subset B$, so

$$B \cap Y = \bigcup_{x\in B\cap Y} (B_X(x, r_x) \cap Y) = \bigcup_{x\in B\cap Y} B_Y(x, r_x).$$

This implies that the restriction of B to Y is open in Y.

Example. For $n \in \mathbb{Z}^+$ define the **sphere** S^n to be the subspace of the Euclidean space $\mathbb{R}^{n+1}$ consisting of all points of distance 1 to the origin:

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

Homeomorphism

A map between two topological spaces is a **homeomorphism** [1] if it is bijective, is continuous, and its inverse map is also continuous.

Two spaces are **homeomorphic** [2] if there is a homeomorphism from one to the other.

[1] phép đồng phôi
[2] đồng phôi

Suppose $f : X \to Y$ is a homeomorphism. For any subset of O of X we have $\left(f^{-1}\right)^{-1}(O) = f(O)$, thus if O is open then $f(O)$ is open, and if O is closed then $f(O)$ is closed. Thus a homeomorphism brings an open set onto an open set, a closed set onto a closed set.

Proposition. *A homeomorphism between two spaces induces a bijection between the two topologies.*

Proof. A homeomorphism $f : (X, \tau_X) \to (Y, \tau_Y)$ induces a map

$$\begin{aligned} \tilde{f} : \tau_X &\to \tau_Y \\ O &\mapsto f(O), \end{aligned}$$

which is a bijection. □

Roughly speaking, in Topology when two spaces are homeomorphic they are considered the same. For example, a "topological sphere" means a topological space which is homeomorphic to a sphere.

3.3 Example. ***Any two balls in a normed space are homeomorphic*** via translations [1] and dilations (scaling) [2]:

$$\begin{aligned} B(0,1) &\to B(a,r) \\ x &\mapsto rx + a. \end{aligned}$$

3.4 Example. ***Any ball in a normed space is homeomorphic to the whole space.*** Consider a map from the unit ball $B(0,1)$ to the whole space in the form $x \mapsto y = f(x) = \varphi(\|x\|)\frac{x}{\|x\|}$, where φ is a homeomorphism from $(0,1)$ onto $(0,\infty)$. The inverse map is $y \mapsto x = \varphi^{-1}\left(\|y\|\right)\frac{y}{\|y\|}$. Such a map φ appears in 1.7, such as $\varphi(t) = \frac{t}{\sqrt{1-t^2}}$, or $\varphi(t) = \frac{t}{1-t}$, or $\varphi(t) = -\ln(1-t)$. For example, take $\varphi(t) = \frac{t}{1-t}$, then $y = f(x) = \frac{x}{1-\|x\|}$, and $x = f^{-1}(y) = \frac{y}{1+\|y\|}$.

3.5 Definition. *An* **embedding** *(or* **imbedding***)* [3] *from a topological space X to a topological space Y is a homeomorphism from X to a subspace of Y, i.e. it is a map $f : X \to Y$ such that the restriction $\tilde{f} : X \to f(X)$ is a homeomorphism. If there is an imbedding from X to Y then we say that X can be* **embedded** *in Y.*

Example. With the subspace topology the inclusion map is an embedding.

Example. The Euclidean line $\mathbb{R}$ can be embedded in the Euclidean plane $\mathbb{R}^2$ as a line in the plane.

Example. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous under the Euclidean topology. Then $\mathbb{R}$ can be embedded into the plane as the graph of f.

---

[1] tịnh tiến
[2] co dãn, vị tự
[3] phép nhúng

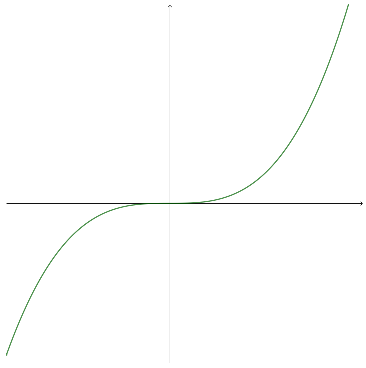


Figure 3.6: The Euclidean real number line can be embedded to the Euclidean plane as the graph of the function $y = x^3$.

3.7 Example (any straight line is homeomorphic to the Euclidean line). A straight line ℓ in a normed space is determined by a point a it passes through and a directional vector $v \neq 0$, consisting of all points x such that $x - a$ is a multiple of v. Thus

$$\ell = \{x \mid x - a = tv, t \in \mathbb{R}\}.$$

Consider the map

$$\begin{aligned} \varphi : \mathbb{R} &\to \ell \\ t &\mapsto x = a + tv. \end{aligned}$$

We have $\|\varphi(s) - \varphi(t)\| = |s - t| \, \|v\|$. This implies that φ is bijective, continuous, and the inverse map is also continuous.

Example. *The sphere S^n minus one point is homeomorphic to the Euclidean space $\mathbb{R}^n$* via the **stereographic projection** [1], see Fig. 3.8.

Specifically, on the sphere S^n minus the North Pole $N = (0, 0, \ldots, 0, 1)$, each point $x = (x_1, x_2, \ldots, x_{n+1})$ corresponds to the intersection $y = (y_1, y_2, \ldots, y_n, 0)$ between the straight line from N to x with the hyperplane $x_{n+1} = 0$. An intersection equation between the line and the hyperplane is:

$$(1 - t)N + tx = y,$$

giving $t = \frac{1}{1-x_{n+1}}$. We get a formula for the stereographic projection:

$$\begin{aligned} S^n \setminus \{(0, 0, \ldots, 0, 1)\} &\to \mathbb{R}^n \times \{0\} \\ (x_1, x_2, \ldots, x_{n+1}) &\mapsto (y_1, y_2, \ldots, y_n, 0), y_i = \frac{1}{1 - x_{n+1}} x_i. \end{aligned}$$

[1] phép chiếu nổi

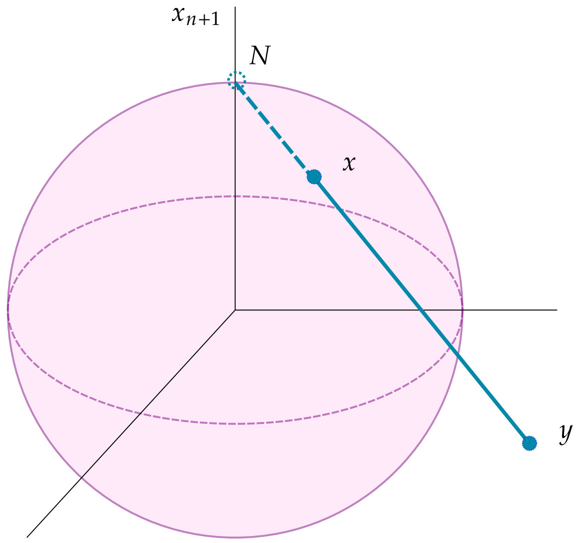


Figure 3.8: The stereographic projection.

We can deduce the inverse map:

$$x_i = \frac{2y_i}{1+\sum_{i=1}^n y_i^2}, 1 \le i \le n, x_{n+1} = \frac{-1+\sum_{i=1}^n y_i^2}{1+\sum_{i=1}^n y_i^2}.$$

Both maps are continuous thus the the stereographic projection is a homeomorphism.

We can also say that the n-dimensional Euclidean space can be embedded onto the n-sphere minus one point.

Problems

3.9. ✓ A map is continuous if and only if the inverse image of a closed set is a closed set.

3.10. ✓ If $f : X \to Y$ and $g : Y \to Z$ are continuous then $g \circ f$ is continuous.

3.11. Prove Proposition 3.1.

3.12. Define an **open map** to be a map such that the image of an open set is an open set. A **closed map** is a map such that the image of a closed set is a closed set.

(a) Show that a homeomorphism is an open map and is also a closed map.

(b) Show that a continuous bijection is a homeomorphism if and only if it is an open map.

3.13. Show that (X, PPX_p) and (X, PPX_q) (see 2.9) are homeomorphic.

3.14. Suppose that X is a normed space. Prove that the topology generated by the norm is exactly the coarsest topology on X such that the norm and the translations (maps of the form $x \mapsto x + a$) are continuous.

3.15. ✓ Show that a subset of a subspace Y of X is closed in Y if and only if it is a restriction of a closed set in X to Y.

3.16. The set $\{x \in \mathbb{Q} \mid -\sqrt{2} < x < \sqrt{2}\}$ is both closed and open in $\mathbb{Q}$ under the Euclidean topology of $\mathbb{R}$.

3.17. ✓ Suppose that X is a topological space and $Z \subset Y \subset X$. Then the relative topology of Z with respect to Y is the same as the relative topology of Z with respect to X.

3.18. ✓ Let X and Y be topological spaces and let $f : X \to Y$.

(a) If Z is a subspace of X, denote by $f|_Z$ the restriction of f to Z. Show that if f is continuous then $f|_Z$ is continuous.

(b) If Z is a space containing Y as a subspace, denote by $\tilde{f}$ be the extension $\tilde{f} : X \to Z$, $\tilde{f}(x) = f(x)$. Show that f is continuous if and only if $\tilde{f}$ is continuous.

Because of these results, later we often extend the target set or restrict the domain set of a continuous function implicitly, without introducing new notations.

3.19. ✓ If $f : X \to Y$ is a homeomorphism and $Z \subset X$ then $f|_Z : Z \to f(Z)$ is a homeomorphism.

3.20. In the Euclidean plane an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is homeomorphic to a circle, and an elliptical region $\frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1$ is homeomorphic to a circular region.

3.21. In the Euclidean plane the upper half-plane $\{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ is homeomorphic to the whole plane.

3.22. Is it true that any two balls in a metric space are homeomorphic?

3.23. Show that any two finite-dimensional normed spaces of the same dimensions over the same fields are homeomorphic.

3.24. On the Euclidean plane $\mathbb{R}^2$, show that:

(a) $\mathbb{R}^2 \setminus \{(0, 0)\}$ and $\mathbb{R}^2 \setminus \{(1, 1)\}$ are homeomorphic.

(b) $\mathbb{R}^2 \setminus \{(0, 0), (1, 1)\}$ and $\mathbb{R}^2 \setminus \{(1, 0), (0, 1)\}$ are homeomorphic.

3.25. Show that $\mathbb{N}$ and $\mathbb{Z}$ are homeomorphic under the Euclidean topology. More generally, prove that any two set-equivalent discrete spaces are homeomorphic.

3.26. Among the spaces $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, each with the Euclidean topology, which one is homeomorphic to another?

3.27. Show that $\mathbb{R}$ with the finite complement topology and $\mathbb{R}^2$ with the finite complement topology are homeomorphic.

3.28. Check that, under the Euclidean topology of $\mathbb{R}^2$, the usual parametrization of the circle using angle, the map $\varphi : [0, 2\pi) \to S^1$ given by $t \mapsto (\cos t, \sin t)$, is a bijective and continuous, but is not a homeomorphism.

3.29 (gluing continuous functions). ✓ Let $X = A \cup B$ where A and B are both open or are both closed in X. Suppose $f : X \to Y$, and $f|_A$ and $f|_B$ are both continuous. Then f is continuous.

Another way to phrase the above statement is the following. Let $g : A \to Y$ and $h : B \to Y$ be continuous and $g(x) = h(x)$ on $A \cap B$. Define

$$f(x) = \begin{cases} g(x), & x \in A \\ h(x), & x \in B. \end{cases}$$

Then f is continuous.

Is it still true if the restriction that A and B are both open or are both closed in X is removed?

3.30. On the Euclidean plane $\mathbb{R}^2$, show that the subspace $(\{0\} \times [0,1]) \cup ([0,1] \times \{0\})$ is homeomorphic to the subspace $\{0\} \times [0,1]$.

3.31. Show that on the Euclidean plane any two triangles are homeomorphic.

Here, given three points A, B, C not on the same straight line, called the vertices, the triangle ABC is the union of the three line segments AB, BC, CA, called the edges.

Can this result be generalized to polygons?

3.32. Prove that any two triangular regions in the Euclidean plane are homeomorphic.

Here a triangular region is the subset of the plane bounded by a triangle, but it might be more precisely and more conveniently to consider a triangular region as the convex hull of three points which are not on a straight line.

3.33. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be any two norms on $\mathbb{R}^n$. Consider the map

$$\begin{aligned} (\mathbb{R}^n, \|\cdot\|_1) &\to (\mathbb{R}^n, \|\cdot\|_2) \\ x &\mapsto \begin{cases} \|x\|_1 \frac{x}{\|x\|_2}, & x \neq 0 \\ 0, & x = 0. \end{cases} \end{aligned}$$

(a) Check that this map is a homeomorphism from $(\mathbb{R}^n, \|\cdot\|_1)$ to $(\mathbb{R}^n, \|\cdot\|_2)$.

(b) Deduce that ***balls with respect to different norms in $\mathbb{R}^n$ are homeomorphic***.

(c) Deduce that in Euclidean $\mathbb{R}^n$ an n-dimensional ball is homeomorphic to an n-dimensional rectangle.

(d) Deduce that in Euclidean $\mathbb{R}^2$ a square (a certain union of four line segments) is homeomorphic to a circle (the set of points of the same distance to a given point).

For extensions of this problem, see 10.19 and 17.22.

3.34. Show that any homeomorphism from S^{n-1} onto S^{n-1} can be extended to a homeomorphism from the unit disk $D^n = B'(0,1)$ onto D^n.

3.35. A space is **homogeneous** [1] if given two points there exists a homeomorphism from the space to itself bringing one point to the other point.

[1] đồng nhất

(a) Show that the sphere S^2 is homogeneous.

(b) Show that being homogeneous is a **topological property**, meaning that if two spaces are homeomorphic and one space is homogeneous then the other space is also homogeneous.

3.36. On the set $\mathbb{R}$ with the finite complement topology, any polynomial is continuous, but a trigonometric function such as the sine function is not continuous.

3.37. An **isometry** or a **distance-preserving map** [1] from a metric space X to a metric space Y is a surjective map $f : X \to Y$ that preserves distance, that is $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$. If there exists such an isometry then X is said to be **isometric** to Y.

(a) In the Euclidean $\mathbb{R}^n$, show that translations, and orthogonal transformations (linear maps of the form $x \mapsto A \cdot x$ where A is an $n \times n$ matrix such that $A^t \cdot A = I_n$) are isometries

For related discussions, see 21.5.

(a) Show that an isometry is a homeomorphism.

(b) Show that being isometric is an equivalence relation among metric spaces.

3.38. * Show that $(\mathbb{R}^2, \|\cdot\|_\infty)$ and $(\mathbb{R}^2, \|\cdot\|_1)$ are isometric, but they are not isometric to $(\mathbb{R}^2, \|\cdot\|_2)$, although the three spaces are homeomorphic. Here $\|(x, y)\|_1 = |x| + |y|$, $\|(x, y)\|_2 = (x^2 + y^2)^{1/2}$, $\|(x, y)\|_\infty = \max\{|x|, |y|\}$.

For generalization to higher dimensions one may use the Mazur–Ulam theorem [P. Lax, *Functional Analysis*, 2002, p. 49].

[1] phép đẳng cấu metric, phép đẳng cấu hình học, hay ánh xạ bảo toàn khoảng cách, ánh xạ đẳng cự

4 Connectedness

Path-connected spaces

Shortly, a space is path-connected if for any two points there is a path connecting them.

Definition. *A* **path** [1] *in a topological space X from a point x to a point y is a continuous map $\alpha : [a,b] \to X$ such that $\alpha(a) = x$ and $\alpha(b) = y$, where the interval of real numbers $[a,b]$ has the Euclidean topology. The space X is said to be* **path-connected** [2] *if for any two points x and y in X there is a path in X from x to y.*

Example. Any normed space is path-connected, and so is any convex subspace of a normed space: any two points x and y are connected by a straight line segment $x + t(y - x)$, $t \in [0,1]$.

Proposition. *The relation on a topological space X whereas a point x is related to a point y if there is a path in X from x to y is an equivalence relation.*

An equivalence class under the above equivalence relation is called a **path-connected component**.

Proof. Every point is path-connected to itself via a constant path.

If α is a path defined on $[a,b]$ then there is a path β defined on $[0,1]$ with the same images (also called the traces of the paths): we can just use the linear homeomorphism $(1-t)a + tb$ from $[0,1]$ to $[a,b]$ and let $\beta(t) = \alpha((1-t)a + tb)$. For convenience we can assume that the domains of paths is the interval $[0,1]$.

If there is a path $\alpha : [a,b] \to X$ from x to y then there is a path from y to x, for examples $\beta : [a,b] \to X$, $\beta(t) = \alpha(a + b - t)$, or $\beta : [b, b+a] \to X$, $\beta(t) = \alpha\left(-\frac{b-a}{a}(t-b) + b\right)$ (just take a linear map bringing the point (b,b) to the point $(b+a, a)$).

If $\alpha : [a,b] \to X$ is a path from x to y and $\beta : [c,d] \to X$ is a path from y to z then there is a path γ from x to z. For example, we may take $\beta_1 : [b, b+(d-c)] \to X$, $\beta_1(t) = \beta(t - b + c)$, then take $\gamma : [a, b+(d-c)] \to X$,

$$\gamma(t) = \begin{cases} \alpha(t), & a \le t \le b, \\ \beta_1(t), & b \le t \le b + (d-c) \end{cases}$$

to go continuously from x to y then to z. ◻

Example. In a normed space, the sphere $S = \{x \mid \|x\| = 1\}$ is path-connected. One way to show this is as follow. If two points x and y are not opposite then they can be connected by the arc $\frac{x+t(y-x)}{\|x+t(y-x)\|}$, $t \in [0,1]$. If x and y are opposite,

[1] đường đi

[2] liên thông đường

we can take a third point z, then take a path from x to z and a path from z to y to get a path from x to y as above. See Fig. 4.1.

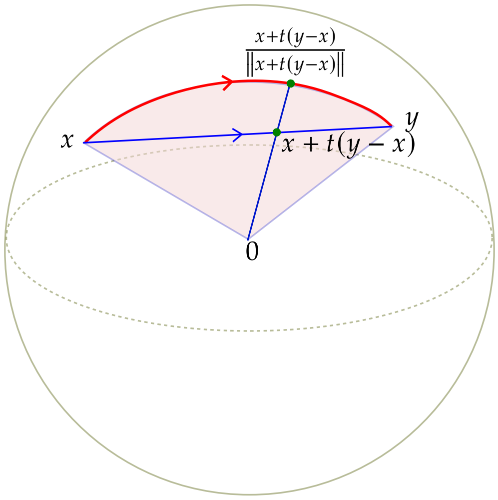


Figure 4.1: A path connecting two points on the sphere.

Connected spaces

Intuitively we say a space is connected to indicate that it has "one piece", whereas a space is disconnected when it has "several pieces".

Definition. *A space is said to be not connected, or* **disconnected***, if it is the union of two non-empty disjoint open subsets. In symbols,* (X,τ) *is disconnected if*

$$\exists U \in \tau, \exists V \in \tau, U \neq \emptyset, V \neq \emptyset, U \cap V = \emptyset, X = U \cup V.$$

Example. The Euclidean real number line minus a point is not connected. For $x \in \mathbb{R}$, $\mathbb{R} \setminus \{x\} = (-\infty, x) \cup (x, \infty)$.

Example. The Euclidean plane minus a straight line is not connected. If the straight line ℓ has equation $f(x,y) = ax + by = c$, then

$$\begin{aligned}\mathbb{R}^2 \setminus \ell &= \{(x,y) \in \mathbb{R}^2 \mid f(x,y) < 0\} \cup \{(x,y) \in \mathbb{R}^2 \mid f(x,y) > c\}\\ &= f^{-1}(-\infty, c) \cup f^{-1}(c, \infty).\end{aligned}$$

Example. The Euclidean plane minus a circle is not connected.

Definition. *A space is said to be* **connected** [1] *if it is not disconnected, equivalently, it is* ***not the union of two non-empty disjoint open subsets****. In symbols,* (X,τ)

[1] liên thông

is connected if[1]

$$\forall U \in \tau, \forall V \in \tau, (U \cap V = \emptyset, X = U \cup V) \implies (U = \emptyset \vee V = \emptyset).$$

Example. A space containing only one point is connected.

We immediately deduce a convenient characterization of connectedness: A topological space is connected if and only if the only subsets which are both closed and open are the empty set and the space itself. So ***connectedness is the non-existence of a non-empty proper subset which is both closed and open***. In a connected space, aside from the trivial sets (the empty set and the space itself) being "open" and being "closed" are mutually exclusive.

When we say that a subset of a topological space is connected we mean that the subset under the subspace topology is a connected space.

Proposition (continuous image of connected space is connected). *If $f : X \to Y$ is continuous and X is connected then $f(X)$ is connected.*

With the above different ways to characterize connectedness, it can be helpful exercises for the readers to write different arguments for some propositions and problems in this section.

Proof. Suppose that U and V are non-empty disjoint open subset of $f(X)$ and $U \cup V = f(X)$. Since $f : X \to f(X)$ is continuous (see 3.18), $f^{-1}(U)$ and $f^{-1}(V)$ are open in X, are non-empty and disjoint, and $f^{-1}(U) \cup f^{-1}(V) = X$. This contradicts the connectedness of X. □

As a consequence connectedness is preserved under homeomorphisms, thus ***connectedness is a topological property***.

The following result is a demonstration of the use of connectedness assumption:

4.2 Proposition (locally constant map on a connected space must be constant). *Let X be connected, and let $f : X \to Y$ be **locally constant**, that is, each point of X has a neighborhood on which f is constant. Then f is constant on X.*

Proof. Fix $x_0 \in X$ and let $c = f(x_0)$. Let $U = \{x \in X \mid f(x) = c\}$, then U is not empty. We shall check that U is both open and closed in X, therefore $U = X$ since X is connected, and so $f = c$ on X.

Given $x_1 \in X$ then since f is locally constant at x_1 there is an open neighborhood V of x_1 such that $f|_V = f(x_1)$. If $x_1 \in U$ then $f(x_1) = c$, therefore $f|_V = c$, implying $V \subset U$. This shows that U is open in X. Similarly if $x_1 \in X \setminus U$ then $f(x_1) \neq c$ therefore for all $x \in V$ we have $f(x) \neq c$, implying $V \subset X \setminus U$. So $X \setminus U$ is also open, therefore U is closed in X. □

[1]Recall that the statement $p \implies q$ is equivalent to the statement $\neg p \vee q$, where $\neg$ denotes negation and $\vee$ denotes "or". Thus $\neg(p \implies q)$ is equivalent to $p \wedge (\neg q)$, where $\wedge$ denotes "and".

Example. Let us prove the following result in Calculus: Let $D \subset \mathbb{R}^n$ be open, connected, let $f : D \subset \mathbb{R}^n \to \mathbb{R}$, and suppose all of the partial derivatives of f are 0, then f must be constant. The case $n = 1$ is included in many Calculus textbooks.

Consider a ball $B(x, r) \subset D$. For any $y \in B(x, r)$, let $g(t) = f(x + t(y - x))$, $t \in [0, 1]$, by the Lagrange Mean value theorem, for some $\theta \in (0, 1)$, we have

$$f(y) - f(x) = g(1) - g(0) = g'(\theta) = \nabla f(x + \theta(y - x)) \cdot (y - x) = 0.$$

Thus f is constant on $B(x, r)$. This implies that f is locally constant on D. Since D is connected we conclude that f is constant on D, by using 4.2 or by repeating the argument there.

The following result is used often in checking that a space is connected:

4.3 Proposition. *If a collection of connected subspaces of a space has non-empty intersection then its union is connected.*

Proof. Consider a topological space and let F be a collection of connected subspaces whose intersection is non-empty. Let A be the union of the collection, $A = \bigcup_{D \in F} D$. Suppose that C is a subset of A that is both open and closed in A. If $C \neq \emptyset$ then there is $D \in F$ such that $C \cap D \neq \emptyset$. Then $C \cap D$ is a subset of D and is both open and closed in D (we are using 3.17 here). Since D is connected and $C \cap D \neq \emptyset$, we must have $C \cap D = D$. This implies C contains the intersection of F. Therefore $C \cap D \neq \emptyset$ for all $D \in F$. The argument above shows that C contains all D in F, that is, $C = A$. We conclude that A is connected. □

4.4 Proposition. *A connected subspace with a limit point added is still connected. Consequently the closure of a connected subspace is connected.*

Proof. Let A be a connected subspace of a space X and let $a \notin A$ be a limit point of A, we show that $A \cup \{a\}$ is connected. Suppose that $A \cup \{a\} = U \cup V$ where U and V are non-empty disjoint open subsets of $A \cup \{a\}$. Suppose that $a \in U$. Then $a \notin V$, so $V \subset A$. Since a is a limit point of A, $U \cap A$ is non-empty. Then $U \cap A$ and V are open subsets of A, by 3.17, which are non-empty and disjoint, while $A = (U \cap A) \cup V$. This contradicts the assumption that A is connected. □

Connected sets in the Euclidean real number line

We determine connected subsets of the set of real numbers under Euclidean topology:

Theorem. *A subspace of the Euclidean real number line is connected if and only if it is an interval.*

Proof. Let us prove that $\mathbb{R}$ is connected. Suppose that $\mathbb{R}$ contains a non-empty, proper, open and closed subset C. Let $x \notin C$.

Suppose that $D = C \cap (-\infty, x) = C \cap (-\infty, x]$ is non-empty. Then D is both open and closed in $\mathbb{R}$, and is bounded from above. Consider $s = \sup D$. Since D is closed and s is a contact point of D, $s \in D$. Since D is open s must belong to an open interval contained in D. But then there are numbers in D which are bigger than s, a contradiction.

If $D = \emptyset$, let $E = C \cap (x, \infty)$, consider $\inf E$ and proceed similarly. Thus we have proved that $\mathbb{R}$ is connected.

For other intervals, we can argue similarly, or we can deduce from the above case as follows. Since $\mathbb{R}$ is connected, by homeomorphisms any intervals (a, b) is connected. By 4.4, since $[a, b]$ is the closure of (a, b), and $(a, b] = (a, b) \cup [\frac{a+b}{2}, b]$, these intervals are also connected.

In the reverse direction, suppose that a subset A of $\mathbb{R}$ is connected. Suppose that $x, y \in A$ and $x < y$. If $x < z < y$ we must have $z \in A$, otherwise the set $\{a \in A \mid a < z\} = \{a \in A \mid a \leq z\}$ will be both closed and open in A. Thus A contains the whole interval $[x, y]$.

Let $a = \inf A$ if A is bounded from below and $a = -\infty$ otherwise. Similarly let $b = \sup A$ if A is bounded from above and $b = \infty$ otherwise. Suppose that A contains more than one element, so that $a < b$. There are sequences $(a_n)_{n \in \mathbb{Z}^+}$ and $(b_n)_{n \in \mathbb{Z}^+}$ of elements in A such that $a < a_n < b_n < b$, and $a_n \to a$ while $b_n \to b$. By the above argument, $[a_n, b_n] \subset A$ for all n. So $(a, b) \subset \bigcup_{n=1}^{\infty} [a_n, b_n] \subset A \subset [a, b]$. It follows that A is either (a, b) or $[a, b)$ or $(a, b]$ or $[a, b]$. ☐

4.5 Example (normed spaces are connected). Since a normed space over $\mathbb{R}$ is the union of all lines passing through the origin, and each line is connected (3.7), it is connected. Therefore a non-empty proper subset of a normed space cannot be both open and closed, in other words, ***in a normed space being closed and being open are exclusive except the two trivial cases***.

4.6 Theorem (Intermediate value theorem). *If X is a connected space and $f : X \to \mathbb{R}$ is continuous, where $\mathbb{R}$ has the Euclidean topology, then the image $f(X)$ is an interval.*

Example. A consequence is the following familiar theorem in Calculus: Let $f : [a, b] \to \mathbb{R}$ be continuous under the Euclidean topology, if $f(a)$ and $f(b)$ have opposite signs then the equation $f(x) = 0$ has a solution.

Theorem (Borsuk–Ulam theorem). *For any continuous real function on the sphere S^n there must be antipodal points (i.e. opposite points via the center of the sphere) where the values of the function are same.* [1]

Proof. Let $f : S^n \to \mathbb{R}$ be continuous. Let $g(x) = f(x) - f(-x)$. Then g is continuous and $g(-x) = -g(x)$. Since S^n is connected (see 4.19), the range $g(S^n)$ is a connected subset of the Euclidean $\mathbb{R}$, and so it is an interval, containing the interval between $g(x)$ and $g(-x) = -g(x)$. Therefore 0 is in the range of g, so there is $x_0 \in S^n$ such that $f(x_0) = f(-x_0)$. □

Connected component

Proposition. *The relation on a topological space X whereas two points are related if both belong to a connected subspace of X is an equivalence relation.*

Under this equivalence relation, the equivalence class containing a point x is equal to the union of all connected subspaces containing x, thus it is the largest connected subspace containing x.

Proof. This relation is an equivalence relation by 4.3.

Consider the equivalence class $[x]$ represented by a point x. By definition, $y \in [x]$ if and only if there is a connected set O_y containing both x and y. Since $O_y \subset [x]$, we have $[x] = \bigcup_{y\in[x]} O_y$. By 4.3, $[a]$ is connected. □

Under the above equivalence relation, the equivalence classes are called the **connected components** of the space. Thus any space is a disjoint union of its connected components. A connected space is a space having only one connected component.

Example. Consider the space $\mathbb{Z}$ with the Euclidean topology inherited from $\mathbb{R}$. This topology is actually the discrete topology on $\mathbb{Z}$, therefore a subset containing only one point is open, and any subset containing more than one point is not connected. Hence the subsets containing only one point are the connected components of $\mathbb{Z}$.

Example. Consider the space $\mathbb{Q}$ with the Euclidean topology inherited from $\mathbb{R}$. If a subset A of $\mathbb{Q}$ contains two elements a and b, $a < b$, then there is an irrational number c between a and b, and $A = ((-\infty, c) \cap A) \cap ((c, \infty) \cap A)$, a union of two open disjoint non-empty subsets, thus A is not connected. Hence the subsets containing only one point are the connected components of $\mathbb{Q}$.

4.7 Proposition. *A map defined on a space with finitely many connected components*

[1] On the surface of the Earth at any moment there two opposite places where temperatures are same!

is continuous if and only if the map is continuous on each component.

This is Problem 4.28.

From this result when study a topological space with finitely many connected components we often only need to work on each connected component, or we may assume the space itself is connected.

Theorem. *If two spaces are homeomorphic then there is a bijection between the collections of connected components of the two spaces.*

Proof. Let $f : X \to Y$ be a homeomorphism. Set

$$\begin{aligned} \tilde{f} : \{[x] \mid x \in X\} &\to \{[y] \mid y \in Y\} \\ [x] &\mapsto [f(x)]. \end{aligned}$$

We check that $\tilde{f}$ is a well-defined map. Suppose that $y \in [x]$. Since f is continuous, the image $f([x])$ is connected. Thus $f(y)$ and $f(x)$ belong to a connected subspace $f([x])$, so $[f(y)] = [f(x)]$.

The map $\tilde{f}$ has an inverse map

$$\begin{aligned} \widetilde{f^{-1}} : \{[y] \mid y \in Y\} &\to \{[x] \mid x \in X\} \\ [y] &\mapsto [f^{-1}(y)]. \end{aligned}$$

Therefore we have a bijection between the two sets. □

From this result we say that ***connectedness is a topological property***. We also say that ***the number of connected components is a topological invariant***. If two spaces have different numbers of connected components then they must be different, that is, not homeomorphic.

Example (a line is not homeomorphic to a plane). Suppose that $\mathbb{R}$ and $\mathbb{R}^2$ under the Euclidean topologies are homeomorphic via a homeomorphism f. Delete any point x from $\mathbb{R}$. By 3.19 the subspaces $\mathbb{R} \setminus \{x\}$ and $\mathbb{R}^2 \setminus \{f(x)\}$ are homeomorphic. But $\mathbb{R} \setminus \{x\}$ is not connected while $\mathbb{R}^2 \setminus \{f(x)\}$ is connected (see 4.14), a contradiction. Thus $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic.

Example (a circle is not homeomorphic to a line). Suppose that there is a homeomorphism f from the Euclidean line $\mathbb{R}$ to the circle S^1. Let $x \in \mathbb{R}$. Then $\mathbb{R} \setminus \{x\}$ is homeomorphic to $S^1 \setminus \{f(x)\}$. But they have different number of connected components, a contradiction.

Relations between path-connectedness and connectedness

Theorem (path-connected $\implies$ connected). *Any path-connected space is connected.*

Proof. This is a consequence of the fact that an interval on the Euclidean real number line is connected. Let X be path-connected. Let $x, y \in X$. There is a path from x to y. The image of this path is a connected subspace of X. That means every point y belongs to the connected component containing x. Therefore X has only one connected component. □

Example (normed spaces are connected). Any normed space is path-connected hence is connected.

On the reverse direction we have:

4.8 Proposition. *If a space is connected and every point has a path-connected neighborhood then it is path-connected.*

Proof. Suppose that X is connected and every point of X has a path-connected neighborhood. Let C be a path-connected component of X. If $x \in X$ is a contact point of C then there is a path-connected neighborhood U in X of x such that $U \cap C \neq \emptyset$. By 4.11, $U \cup C$ is path-connected, so $U \cup C \subset C$, thus $U \subset C$. This implies that every contact point of C is an interior point of C, thus C is both open and closed in X. Since X is connected, $C = X$. □

A topological space is said to be **locally path-connected** if every neighborhood of a point contains an open path-connected neighborhood of that point.

Corollary. *A connected, locally path-connected space is path-connected.*

Example. Open subsets of normed spaces are locally path-connected, because balls are path-connected, hence ***for open subsets of normed spaces connectedness and path-connectedness are same***.

Proposition (connected $\neq$ path-connected). *There exists a space which is connected but is not path-connected.*

Proof. The closure in the Euclidean plane of the graph of the function $y = \sin\frac{1}{x}$, $x > 0$, is often called the **Topologist's sine curve**. This is an example of a space which is connected but is not path-connected.

Denote $A = \{(x, \sin\frac{1}{x}) \mid x > 0\}$ and $B = \{0\} \times [-1, 1]$. Then the Topologist's sine curve is $X = A \cup B$.

By 4.20 the set A is connected. Each point of B is a limit point of A, so by 4.4 X is connected.

Suppose that there is a path $\gamma(t) = (x(t), y(t))$, $t \in [0, 1]$ from the origin $(0, 0)$ on B to a point on A, we show that there is a contradiction.

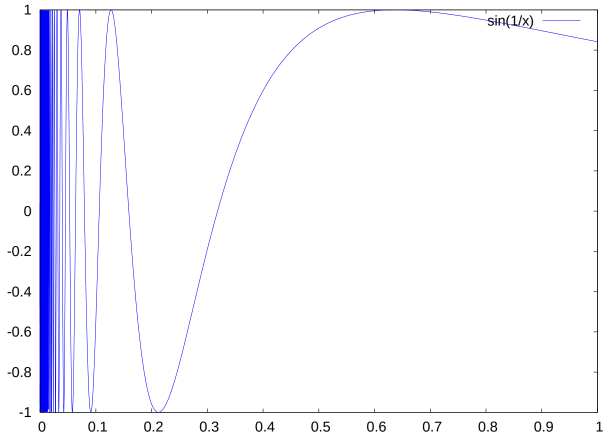


Figure 4.9: The topologist's sine curve.

Let $t_0 = \sup\{t \in [0,1] \mid x(t) = 0\}$. Then $x(t_0) = 0$, $t_0 < 1$, and $x(t) > 0$ for all $t > t_0$. Thus t_0 is the moment when the path γ departs from B. We can see that the path jumps immediately when it departs from B. Thus we will show that $y(t)$ cannot be continuous at t_0 by showing that for any $\delta > 0$ there are $t_1, t_2 \in (t_0, t_0 + \delta)$ such that $y(t_1) = 1$ and $y(t_2) = -1$.

To find t_1, note that the set $x([t_0, t_0 + \frac{\delta}{2}])$ is an interval $[0, x_0]$ where $x_0 > 0$. There exists an $x_1 \in (0, x_0)$ such that $\sin \frac{1}{x_1} = 1$: we just need to take $x_1 = \frac{1}{\frac{\pi}{2} + k2\pi}$ with sufficiently large k. There is $t_1 \in (t_0, t_0 + \frac{\delta}{2}]$ such that $x(t_1) = x_1$. Then $y(t_1) = \sin \frac{1}{x(t_1)} = 1$. We can find t_2 similarly. □

Problems

4.10. If $f : X \to Y$ is continuous and X is path-connected then $f(X)$ is path-connected.

4.11. If a collection of path-connected subspaces of a space has non-empty intersection then its union is path-connected.

4.12. The path-connected component containing a point x is the union of all path-connected subspaces containing x, thus it is the largest path-connected subspace containing x.

4.13. If two space are homeomorphic then there is a bijection between the collections of path-connected components of the two spaces. In particular, if one space is path-connected then the other space is also path-connected.

4.14. Show that the Euclidean plane with countably many points removed is path-connected.

4.15. Is the Euclidean $\mathbb{R}^3$ minus countably many straight lines path-connected?

4.16. A topological space is locally path-connected if and only if the collection of all open path-connected subsets is a basis for the topology.

4.17. A space is connected if and only if whenever it is a union of two non-empty disjoint subsets at least one set must contain a contact point of the other set.

4.18. Here is a different proof that any interval of real numbers is connected. Suppose that A and B are non-empty, disjoint subsets of $(0,1)$ whose union is $(0,1)$. Let $a \in A$ and $b \in B$. Let $a_0 = a$, $b_0 = b$, and for each $n \geq 1$ consider the middle point of the segment from a_n to b_n. If $\frac{a_n+b_n}{2} \in A$ then let $a_{n+1} = \frac{a_n+b_n}{2}$ and $b_{n+1} = b_n$; otherwise let $a_{n+1} = a_n$ and $b_{n+1} = \frac{a_n+b_n}{2}$. Then:

(a) The sequence $(a_n)_{n\geq 1}$ is a Cauchy sequence, hence is convergent to a number c.

(b) The sequence $(b_n)_{n\geq 1}$ is also convergent to c. This implies that $(0,1)$ is connected.

4.19. Find several ways to show that the sphere S^n is connected and is path-connected.

4.20. ✓ If $f : \mathbb{R} \to \mathbb{R}$ is continuous under the Euclidean topology then its graph is connected in the Euclidean plane. Moreover the graph is homeomorphic to $\mathbb{R}$.

4.21. Let X be a topological space and let A_i, $i \in I$ be connected subspaces. If $A_i \cap A_j \neq \emptyset$ for all $i, j \in I$ then $\bigcup_{i\in I} A_i$ is connected.

4.22. Let X be a topological space and let A_i, $i \in \mathbb{Z}^+$ be connected subsets. If $A_i \cap A_{i+1} \neq \emptyset$ for all $i \geq 1$ then $\bigcup_{i=1}^{\infty} A_i$ is connected.

4.23. Let X be a topological space. A map $f : X \to Y$ is called a **discrete map** if Y has the discrete topology and f is continuous. Show that X is connected if and only if all discrete maps on X are constant.

4.24. What are the connected components of $\mathbb{Q}^2$ under the Euclidean topology?

4.25. Find the connected components of the Cantor set (see 2.18).

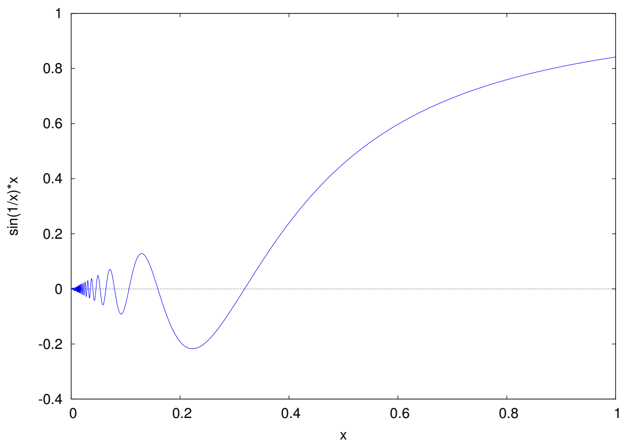
4.26. Show that every connected component is a closed subset.

4.27. Show that if a space has finitely many components then each component is both open and closed.

Is it still true if there are infinitely many components?

4.28. Prove 4.7. Is it still correct if the space has infinitely many connected components?

4.29. Let $X = \{(x, x\sin\frac{1}{x}) \mid x > 0\} \cup \{(0,0)\}$, that is, the graph of the function $x\sin\frac{1}{x}$, $x > 0$, with the origin added, under the Euclidean topology of the plane. Is X connected? Is X path-connected?



4.30. The Topologist's sine curve is not locally path-connected.

4.31. The set of $n \times n$-matrix with real coefficients, denoted by $M(n, \mathbb{R})$, could be naturally considered as a subset of the Euclidean space $\mathbb{R}^{n^2}$ by considering entries of a matrix as coordinates, via the map

$$(a_{i,j}) \longmapsto (a_{1,1}, a_{2,1}, \ldots, a_{n,1}, a_{1,2}, a_{2,2}, \ldots, a_{n,2}, a_{1,3}, \ldots, a_{n-1,n}, a_{n,n}).$$

The **Orthogonal Group** $O(n)$ is defined to be the group of matrices representing orthogonal linear maps of $\mathbb{R}^n$, that is, linear maps that preserve inner product. Thus

$$O(n) = \{A \in M(n, \mathbb{R}) | \ A \cdot A^T = I_n\}.$$

The **Special Orthogonal Group** $SO(n)$ is the subgroup of $O(n)$ consisting of all orthogonal matrices with determinant 1.

(a) Show that any element of SO(2) is of the form

$$R(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

This is a rotation in the plane around the origin with an angle φ. Thus SO(2) is the group of rotations on the plane around the origin.

(b) Show that SO(2) is path-connected.

(c) How many connected components does O(2) have?

(d) * It is known [F. Gantmacher, *Theory of matrices*, vol. 1, Chelsea, 1959, p. 285], that for any matrix $A \in O(n)$ there is a matrix $P \in O(n)$ such that $A = PBP^{-1}$ where B is a matrix of the form

$$\begin{pmatrix} R(\varphi_1) & & & & & \\ & \ddots & & & & \\ & & R(\varphi_k) & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}.$$

Using this fact, prove that $SO(n)$ is path-connected.

4.32. Classify the alphabetical characters up to homeomorphisms, that is, which of the following characters are homeomorphic to each other as subspaces of the Euclidean plane?

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Try to give rigorous arguments. Note that the result depends on the font you use!

Do the same for the Vietnamese alphabetical characters:

A Ă Â B C D Đ E Ê G H I K L M N O Ô Ơ P Q R S T U Ư V X Y

5 Convergence

Sequence

Recall that a sequence in a set X is a map $\mathbb{Z}^+ \to X$, a countably indexed family of elements of X. Given a sequence $x : \mathbb{Z}^+ \to X$ the element $x(n)$ is often denoted as x_n, and the sequence is often denoted by $(x_n)_{n\in\mathbb{Z}^+}$ or $\{x_n\}_{n\in\mathbb{Z}^+}$.

When X is a metric space, the sequence is said to be convergent [1] to x if x_n can be as closed to x as we want provided n is sufficiently large, i.e.

$$\forall \epsilon > 0, \exists N \in \mathbb{Z}^+, \forall n \geq N, d(x_n, x) < \epsilon,$$

which is equivalent to

$$\forall U \in \{U \in \tau_X \mid U \ni x\}, \exists N \in \mathbb{Z}^+, \forall n \geq N, x_n \in U,$$

which can be used in topological spaces.

In metric spaces the topology can be described by sequences. From 2.13, limit points and contact points are limits of sequences. But this is not true in general topological spaces.

5.1 Example (sequence is not adequate for convergence). Consider $\mathbb{R}$ with the countable complement topology. Any point $x \geq 0$ is a contact point of the set $(-\infty, 0)$ since every open set containing x has countable complement therefore cannot be disjoint from $(-\infty, 0)$ which is uncountable. Let $(x_n)_{n\in\mathbb{Z}^+}$ be any sequence in $(-\infty, 0)$. The set $\mathbb{R} \setminus \{x_n \mid n \in \mathbb{Z}^+\}$ is an open set containing x containing no element of the sequence $(x_n)_{n\in\mathbb{Z}^+}$. Thus the sequence $(x_n)_{n\in\mathbb{Z}^+}$ cannot converge to x. Although x is a contact point of the set $(-\infty, 0)$, not sequence in this set can converge to x. In this space the notion of contact point cannot be expressed by the notion of sequence.

In general topological spaces we need to use a notion more general than sequence. Roughly speaking, sequences, having countable indexes, might not be sufficient for discribing contact points, closed subsets, and continuity in general topological spaces, unlike in metric spaces. Arbitrary index is needed.

Net

Definition. *A* **directed set** [2] *is a (partially) ordered set such that for any two indices there is an index greater than or equal to both. In symbols:*

$$\forall i \in I, \forall j \in I, \exists k \in I, k \geq i \wedge k \geq j.$$

[1] hội tụ

[2] tập được định hướng

*A **net** [1], also called a **generalized sequence** [2], in a space is a map from a directed set to that space. A net on a space X with index set a directed set I is a map $x : I \to X$. It is an element of the Cartesian product set $\prod_{i \in I} X$. Writing $x_i = x(i)$ we often denote the net as $(x_i)_{i \in I}$ or $\{x_i\}_{i \in I}$.*

Example. Nets with index set $I = \mathbb{Z}^+$ with the usual order of $\mathbb{Z}^+$ are exactly sequences.

5.2 Example. Let X be a topological space and $x \in X$. Let I be the collection of all open neighborhoods of x. Define an order on I by $U \leq V \iff U \supset V$. Then I becomes a directed set.

Definition. *A net $(x_i)_{i \in I}$ is said to be **convergent** to $x \in X$ if for each neighborhood U of x there is an index $i \in I$ such that if $j \geq i$ then x_j belongs to U. The point x is called a **limit** of the net $(x_i)_{i \in I}$ and we often write $x_i \to x$. In symbols:*

$$\forall U \in \{U \in \tau \mid U \ni x\}, \exists i \in I, \forall j \in \{j \in I \mid j \geq i\}, x_j \in U.$$

Example. Convergence of nets with index set $I = \mathbb{Z}^+$ with the usual order is exactly convergence of sequences. Many statements about convergence in metric spaces could be carried over to topological spaces by simply replacing sequences with nets.

Example. Let $X = \{x_1, x_2, x_3\}$ with topology $\{\emptyset, X, \{x_1, x_3\}, \{x_2, x_3\}, \{x_3\}\}$. The net (x_3) converges to x_1, x_2, x_3. The net (x_1, x_2) converges to x_2.

5.3 Proposition. *A point $x \in X$ is a contact point of a subset $A \subset X$ if and only if there is a net in A convergent to x.*

As a consequence, subset of a space is closed if and only if any limit of any net in that set belongs to that set. A subset is open if and only if no limit of a net outside of that set belong to that set.

This result implies that ***topologies are determined by convergences***.

Proof. ($\Longleftarrow$) Suppose that there is a net $(x_i)_{i \in I}$ in A convergent to x. Let U be an open neighborhood of x. There is an $i \in I$ such that for $j \geq i$ we have $x_j \in U$, in particular $x_i \in U \cap A$. Thus x is a contact point of A.

($\Longrightarrow$) Suppose that x is a contact point of A. Consider the directed set I consisting of all the open neighborhoods of x with the partial order $U \leq V$ if $U \supset V$, see Example 5.2. For any open neighborhood U of x there is an element $x_U \in U \cap A$. For any $V \geq U$, $x_V \in V \subset U$. Thus $(x_U)_{U \in I}$ is a net in A convergent to x. (This construction of the net $(x_U)_{U \in I}$ involves the Axiom of choice.)

[1] lưới

[2] dãy suy rộng

A subset A is closed if and only if it contains the set of its contact points (see 2.11). Thus A is closed if and only if a limit of a net in A belongs to A. And A is open if and only if the complement of A is closed. □

Remark. When can nets be replaced by sequences? By examining the proof of 5.3 we can see that the term net can be replaced by the term sequence if there is a countable collection F of neighborhoods of x such that any neighborhood of x contains a member of F. In this case the point x is said to have a countable **neighborhood basis**. A space having this property at every point is said to be a **first countable space**. A metric space is such a space, where each point has a neighborhood basis consisting of balls of rational radii.

5.4 Proposition. *Let τ_1 and τ_2 be two topologies on X. If convergence in τ_1 implies convergence in τ_2 then τ_1 is finer than τ_2. In symbols: if for all nets x_i and all points x, $x_i \xrightarrow{\tau_1} x \implies x_i \xrightarrow{\tau_2} x$, then $\tau_2 \subset \tau_1$. As a consequence, **if convergences are same then topologies are same**.*

Proof. If convergence in τ_1 implies convergence in τ_2 then contact points in τ_1 are contact points in τ_2. Therefore closed sets in τ_2 are closed sets in τ_1, and so are open sets. □

Example. In $\mathbb{R}^n$, consider the norm $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$, for $p \in \mathbb{R}$, $p \geq 1$. It is immediate that convergence in this norm is equivalent to point-wise convergence (convergence in each coordinate), hence convergence in all p-norms are same, thus the topologies generated by all p-norms are same.

Similarly to the case of metric spaces, we have:

5.5 Proposition. *A function f is continuous at x if and only if for all nets (x_i), $x_i \to x \implies f(x_i) \to f(x)$.*

Proof. The proof is simply a repeat of the proof for the case of metric spaces.

($\implies$) Suppose that f is continuous at x. Let U is a neighborhood of $f(x)$. Then $f^{-1}(U)$ is a neighborhood of x in X. Since (x_i) is convergent to x, there is an $i \in I$ such that for all $j \geq i$ we have $x_j \in f^{-1}(U)$, which implies $f(x_j) \in U$.

($\impliedby$) We will show that if U is an open neighborhood in Y of $f(x)$ then $f^{-1}(U)$ is a neighborhood in X of x. Suppose the contrary, then x is not an interior point of $f^{-1}(U)$, so it is a limit point of $X \setminus f^{-1}(U)$. By 5.3 there is a net (x_i) in $X \setminus f^{-1}(U)$ convergent to x. By the assumption, $f(x_i) \in Y \setminus U$ is convergent to $f(x) \in U$. This contradicts the assumption that U is open. □

Separation

We know any metric on a set induces a topology on that set. On a topological space, if the topology can be induced from a metric, we say that the topological space is **metrizable** [1].

Example. On any set X, the discrete topology is generated by the metric

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

Indeed, for any $x \in X$, the set containing one point $\{x\} = B(x,1)$ is open, therefore any subset of X is open.

On the other hand, if X has more than one element then no metric can generate the trivial topology on X. Indeed, if X has two different elements x and y then the ball $B\left(x, \frac{d(x,y)}{2}\right)$ should be a non-empty open proper subset of X. Thus the trivial topology is not metrizable.

Definition. *Define the following types of topological spaces:*

T_1 *A topological space is called a T_1-space if for any two different points x and y there is an open set containing x but not y and an open set containing y but not x.*

T_2 *A topological space is called a T_2-space or a* **Hausdorff space** [2] *if for any two different points x and y there are disjoint open sets U and V such that $x \in U$ and $y \in V$. Briefly, two different points can be separated by open sets.*

T_3 *A T_1-space is called a T_3-space or a* **regular space** [3] *if for any point x and a closed set F not containing x there are disjoint open sets U and V such that $x \in U$ and $F \subset V$.* [4]*Briefly, a point and a disjoint closed set can be separated by open sets.*

T_4 *A T_1-space is called a T_4-space or a* **normal space** [5] *if for any two disjoint closed sets F and G there are disjoint open sets U and V such that $F \subset U$ and $G \subset V$. Briefly, two disjoint closed sets can be separated by open sets.*

Proposition. *A space is a T_1 space if and only if any subset containing exactly one point is a closed set.*

Proof. If a space X is a T_1 space, given $x \in X$, for any $y \neq x$ there is an open set U_y that does not contain x. Then $X \setminus \{x\} = \bigcup_{y \neq x} U_y$. So $X \setminus \{x\}$ is open. If any point constitutes a closed set, then two points $y \neq x$ can be separated by the open sets $X \setminus \{x\}$ and $X \setminus \{y\}$. □

Proposition **($T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1$).** *If a space is a T_i-space then it is T_{i-1}-space, for* $2 \leq i \leq 4$.

[1] mêtríc hóa được, khả mêtríc

[2] Felix Hausdorff contributed to topology during the first half of the 20th century.

[3] chính tắc

[4] We include T_1 requirement for regular and normal spaces, as in Munkres [Mun00]. Some authors such as Kelley [Kel55] do not include the T_1 requirement.

[5] chuẩn tắc

Proof. The inclusion of a T_1 requirement to T_3 and T_4 ensures that a set containing exactly one point is a closed set. □

Example. Any space with the discrete topology is a normal space.

In particular, a space containing only one point is a normal space.

Example. Any metric space is a Hausdorff space, since if $x \neq y$ then the balls $B\left(x, \frac{d(x,y)}{2}\right)$ and $B\left(y, \frac{d(x,y)}{2}\right)$ separate x and y.

As a consequence, if a space is not a Hausdorff space then it cannot be metrizable.

There is a stronger result:

5.6 Proposition. *Any metric space is a normal space.*

Proof. Given a metric space (X, d), let A and B be two closed disjoint subsets.

Since $X \setminus B$ is open and $A \subset X \setminus B$, for each $x \in A$, there is a ball $B(x, \epsilon_x) \subset X \setminus B$. Let $U = \bigcup_{x \in A} B(x, \frac{\epsilon_x}{2})$, then $A \subset U \subset X \setminus B$.

We check that $\overline{U} \subset X \setminus B$. Let $y \in \overline{U}$. There is a sequence of elements of U convergent to y, hence there are $y_n \in U$ such that $d(y_n, y) < \frac{1}{n}$, for each $n \in \mathbb{Z}^+$. For each n, there is $x_n \in A$ such that $d(y_n, x_n) < \frac{\epsilon_{x_n}}{2}$. Hence $d(y, x_n) < \frac{1}{n} + \frac{\epsilon_{x_n}}{2} < \epsilon_{x_n}$ if n is sufficiently large, which leads to $y \in B(x_n, \epsilon_{x_n}) \subset X \setminus B$.

Let $V = X \setminus \overline{U}$. Then V is open, $B \subset V$, and $U \cap V = \emptyset$. □

5.7 Example. The set of all real numbers under the finite complement topology is a T_1-space but is not a T_2-space.

There are examples of a T_2-space which is not T_3, and a T_3-space which is not T_4, see 5.27, [Mun00, p. 197], [SJ70].

Similarly to the proof of 5.6, the following proposition is a useful characterization of normal spaces:

5.8 Proposition. *A T_1-space is a normal space if and only if given a closed set C and an open set U containing C there is an open set V such that $C \subset V \subset \overline{V} \subset U$.*

Roughly, in a normal space given a closed set inside an open set it is possible to enlarge the closed set while remains inside the open set.

Proof. Suppose that X is normal. Since $X \setminus U$ is closed and disjoint from C there is an open set V containing C and an open set W containing $X \setminus U$ such that V and W are disjoint. Then $V \subset (X \setminus W)$, so $\overline{V} \subset (X \setminus W) \subset U$.

In the reverse direction, given a closed set F and a closed set G disjoint from F, let $U = X \setminus G$. There is an open set V containing F such that $V \subset \overline{V} \subset U$. Then V and $X \setminus \overline{V}$ separate F and G. □

In Analysis, we often consider Hausdorff spaces because of uniqueness of limit:

5.9 Proposition. *A space is a Hausdorff space if and only if any net has at most one limit.*

Proof. Suppose the space is a Hausdorff space. Suppose that a net $(x_i)_{i\in I}$ is convergent to two different limits x and y. Since the space is Hausdorff, there are disjoint open neighborhoods U and V of x and y. There is $i \in I$ such that for $k \geq i$ we have $x_k \in U$, and there is $j \in I$ such that for $k \geq j$ we have $x_k \in U$. There is a $k \in I$ such that $k \geq i$ and $k \geq j$ (the index set is directed), and the point x_k is in $U \cap V$, a contradiction.

Suppose that the space is not a Hausdorff space, then there are two points x and y that could not be separated by open sets. Consider the index set I whose elements are pairs (U, V) of open neighborhoods of x and y, with the order $(U_1, V_1) \leq (U_2, V_2)$ if $U_1 \supset U_2$ and $V_1 \supset V_2$. We can check immediately that with this order the index set is directed. Since $U \cap V \neq \emptyset$, we can take $z_{(U,V)} \in U \cap V$. Then the net $\left(z_{(U,V)}\right)_{(U,V)\in I}$ is convergent to both x and y, contradicting the uniqueness of limit. □

Problems

5.10. Let $I = (0, \infty) \subset \mathbb{R}$. For $i, j \in I$, define $i \leq_I j$ if $i \geq_{\mathbb{R}} j$ (i is less than or is equal to j as indexes if i is greater than or is equal to j as real numbers). On $\mathbb{R}$ with the Euclidean topology, consider the net $(x_i = i)_{i\in I}$. Is this net convergent?

5.11. On $\mathbb{R}$ with the finite complement topology and the usual order of real numbers, consider the net $(x_i = i)_{i\in\mathbb{R}}$. Where does this net converge to?

5.12. Reconsider Problems 2.25 and 2.27 by using 5.4.

5.13. Show that if $f : X \to Y$ is continuous and $A \subset X$ then $f(\overline{A}) \subset \overline{f(A)}$.

(a) By using nets.

(b) By not using nets (by using open sets instead).

5.14. Let Y be a T_1-space, and let $f : X \to Y$ be continuous. Suppose that $A \subset X$ and $f(x) = c$ on A, where c is a constant. Show that $f(x) = c$ on $\overline{A}$.

(a) By using nets.

(b) By not using nets.

5.15. ✓ Let Y be a Hausdorff space and let $f, g : X \to Y$ be continuous. Show that the set $\{x \in X \mid f(x) = g(x)\}$ is closed in X.

(a) By using nets.

(b) By not using nets.

(c) Show that, as a consequence, if f and g agree on a **dense** [1] subspace of X (meaning the closure of that subspace is X) then they agree on X.

[1] dày đặc, trù mật

5.16. * Let $(A, \leq)$ be a well-ordered uncountable set (see 1.26). The smallest element of A is be denoted by 0. If A does not have a biggest element then add an element to A and define that element to be the biggest one, denoted by ∞. For $a, b \in A$ denote $[a, b] = \{x \in A \mid a \leq x \leq b\}$ and $[a, b) = \{x \in A \mid a \leq x < b\}$. Thus we can write $A = [0, \infty]$.

Let Ω be the smallest element of the set $\{a \in A \mid [0, a] \text{ is uncountable}\}$ (this set is non-empty since it contains ∞).

(a) Show that $[0, \Omega)$ is uncountable, and for all $a \in A$, $a < \Omega$ the set $[0, a]$ is countable.

(b) Consider $[0, \Omega]$ with the order topology. Show that Ω is a limit point of $[0, \Omega]$.

(c) Show that every countable subset of $[0, \Omega)$ is bounded in $[0, \Omega)$, therefore a sequence in $[0, \Omega)$ cannot converge to Ω.

5.17 (filter). * A **filter** [1] on a set X is a collection F of non-empty subsets of X such that:

(a) if $A, B \in F$ then $A \cap B \in F$,

(b) if $A \subset B$ and $A \in F$ then $B \in F$.

For example, given a point, the collection of all neighborhoods of that point is a filter.

A filter is said to be convergent to a point if every neighborhood of that point is an element of the filter.

A **filter-base** [2] is a collection G of non-empty subsets of X such that if $A, B \in G$ then there is $C \in G$ such that $C \subset (A \cap B)$.

If G is a filter-base in X then the filter generated by G is defined to be the collection of all subsets of X each containing an element of G: $\{A \subset X \mid \exists B \in G,\ B \subset A\}$.

For example, in a metric space, the collection of all open balls centered at a point is the filter-base for the filter consisting of all neighborhoods of that point.

A filter-base is said to be convergent to a point if the filter generated by it converges to that point.

(a) Show that a filter-base is convergent to x if and only if every neighborhood of x contains an element of the filter-base.

(b) Show that a point $x \in X$ is a limit point of a subset A of X if and only if there is a filter-base of X whose elements are subsets of $A \setminus \{x\}$ convergent to x.

(c) Show that a map $f : X \to Y$ is continuous at x if and only if for any filter-base F which is convergent to x, the filter-base $f(F)$ is convergent to $f(x)$.

Filter gives an alternative to net for describing convergence. For more see [Dug66, p. 209], [Eng89, p. 49], [Kel55, p. 83].

5.18 (Riemann integral by nets). * Here we reformulate the Riemann integral in terms of convergences of nets.

First we recall a setting of Riemann integral via lower sums and upper sums, see for examples [Lan97], [Vgt3].

[1] lọc
[2] cơ sở lọc

Define an n-dimensional rectangle to be a subset of Euclidean space $\mathbb{R}^n$ of the form $I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ with $a_i < b_i$ for all $1 \le i \le n$. The volume of the rectangle I is defined to be the number $|I| = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n)$.

A partition of the interval $[a, b]$ is a finite subset of $[a, b]$ which contains both a and b. We can label the elements of the partition as $x_0, x_1, \ldots, x_m$ with $a = x_0 < x_1 < x_2 < \cdots < x_m = b$. Each interval $[x_{i-1}, x_i]$ is a sub-interval of $[a, b]$ with respect to the partition. A partition of the the rectangle $I = \prod_{i=1}^n [a_i, b_i]$ is a Cartesian product of partitions of $[a_i, b_i]$, namely if P_i is a partition of $[a_i, b_i]$ then $P = \prod_{i=1}^n P_i$ is a partition of I.

A sub-rectangle with respect to the partition P of a rectangle I is a product $\prod_{i=1}^n T_i$ where T_i is a sub-interval of $[a_i, b_i]$ with respect to the partition P_i. Let $\mathrm{Rec}(P)$ be the set of all sub-rectangles with respect to the partition P.

Let $f : I \to \mathbb{R}$ be bounded. With $P \in \mathrm{Par}(I)$, let

$$L(f, P) = \sum_{R \in \mathrm{Rec}(P)} (\inf_R f)\,|R|\,,$$

as a lower sum and

$$U(f, P) = \sum_{R \in \mathrm{Rec}(P)} (\sup_R f)\,|R|$$

as an upper sum. The function f is called Riemann integrable if

$$\sup_{P \in \mathrm{Par}(I)} L(f, P) = \inf_{P \in \mathrm{Par}(I)} U(f, P)$$

and then that number is called the Riemann integral of f (this is also called the Darboux integral and is known to coincide with the formulation via Riemann sum, see [Vgt3]).

(a) Let $\mathrm{Par}(I)$ be the set of all partitions of I. On $\mathrm{Par}(I)$ we define the following order relation:

$$P \le P' \iff P \subset P'.$$

For two partitions $P = \prod_{i=1}^n P_i$ and $P' = \prod_{i=1}^n P_i'$, let $P'' = \prod_{i=1}^n P_i''$ with $P_i'' = P_i \cup P_i'$. Show that $P \le P''$ and $P' \le P''$. Deduce that $(\mathrm{Par}(I), \le)$ is a directed set.

(b) Consider the lower sums and the upper sums as nets indexed by partitions. Show that f is Riemann integrable if and only if

$$\lim L(f, P) = \lim U(f, P)$$

and then the Riemann integral is equal to that limit.

5.19. If a finite set is a T_1-space then the topology is the discrete topology.

5.20. Is the space (X, PPX_p) (see 2.9) a Hausdorff space?

5.21. ✓ Show that the separation properties T_i, $1 \le i \le 4$, are topological properties, that means, if a space X is T_i-space and is homeomorphic to a space Y, then Y is also a T_i-space.

5.22. Show that in a Hausdorff space three different points can be separated by open sets. Can this statement be generalized?

5.23. ✓ Show that a subspace of a Hausdorff space is a Hausdorff space.

5.24. Show that a closed subspace of a normal space is normal.

5.25. Show that a T_1-space X is regular if and only if given a point x and an open set U containing x there is an open set V such that $x \in V \subset \overline{V} \subset U$.

5.26. Let X be a normal space. Suppose that U_1 and U_2 are open sets in X satisfying $U_1 \cup U_2 = X$. Show that there is an open sets V_1 such that $\overline{V_1} \subset U_1$ and $V_1 \cup U_2 = X$. This means in a normal space given any open cover we can get a cover consists of smaller sets.

5.27 (T_2 but not T_3). Show that the set $\mathbb{R}$ with the topology generated by all the subsets of the form (a, b) and $(a, b) \cap \mathbb{Q}$ is a Hausdorff space but is not a regular space.

5.28 (distance function). Let (X, d) be a metric space. For $x \in X$ and $A \subset X$ we define the distance from x to A to be $d(x, A) = \inf\{d(x, y) \mid y \in A\}$.

(a) Check that $d(x, A)$ is a continuous function with respect to x.

(b) Check that $d(x, A) = 0 \iff x \in \overline{A}$.

5.29 (another proof that every every metric space is a normal space). Let (X, d) be a metric space and let A and B be disjoint closed subsets of X. Let $U = \{x \mid d(x, A) < d(x, B)\}$ and $V = \{x \mid d(x, A) > d(x, B)\}$, see 5.28. Show that $A \subset U$, $B \subset V$, $U \cap V = \emptyset$, and both U and V are open. Deduce 5.6.

5.30. Let (X, d) be a metric space. For two subsets A and B of X, define $d(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\}$.

(a) Check that $d(\{a\}, B) = d(a, B)$, as defined in 5.28.

(b) Check that $d(A, B) = \inf\{d(a, B) \mid a \in A\} = \inf\{d(A, b) \mid b \in B\}$.

(c) When is $d(A, B) = 0$? Is d a metric?

For more, see 5.31.

5.31 (Hausdorff distance). Given a metric space (X, d), the **Hausdorff distance** between two bounded subsets A and B is defined to be $d_H(A, B) = \max\{\sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A)\}$. Show that d_H is a metric on the set of all closed bounded subsets of X.

For more, see [BBI01, ch. 7].

5.32. * Let X be a normal space, let $f : X \to Y$ be a surjective, continuous, and closed map. Prove that Y is a normal space.

6 Compactness

Compact metric spaces

Recall from Analysis that a metric space is said to be compact if every sequence has a convergent subsequence. For more general topological spaces, a notion of compactness in terms of open covers, developed gradually during the early 20th century, has been in common use.

A **cover** [1] of a set X is a collection of subsets of X whose union is X. A subcollection of a cover which is itself a cover is called a **subcover** [2]. A cover is said to be an **open cover** if each member of the cover is an open subset of X.

6.1 Lemma (Lebesgue number). *In a compact metric space, for any open cover there exists a number $\epsilon > 0$ such that any ball of radius ϵ is contained in an element of the cover.*

Proof. Let O be a cover of a sequentially compact metric space X. Suppose the opposite of the conclusion, that is for any number $\epsilon > 0$ there is a ball $B(x, \epsilon)$ not contained in any of the element of O. Take a sequence of such balls $B(x_n, \frac{1}{n})$. The sequence $(x_n)_{n\in\mathbb{Z}^+}$ has a subsequence $(x_{n_k})_{k\in\mathbb{Z}^+}$ converging to x. There is $r > 0$ such that $B(x, r)$ is contained in an element U of O. Take k sufficiently large such that $\frac{1}{n_k} < \frac{r}{2}$ and $x_{n_k} \in B(x, \frac{r}{2})$. Then $B(x_{n_k}, \frac{1}{n_k}) \subset B(x_{n_k}, \frac{r}{2}) \subset B(x, r) \subset U$, a contradiction. □

6.2 Theorem. *A metric space is compact if and only if every open cover has a finite subcover.*

Proof. ($\Leftarrow$) Suppose that the metric space X is not compact. There is a sequence $(x_n)_{n\in\mathbb{Z}^+}$ with no convergent subsequence. For an arbitrary point $x \in X$ there is an open neighborhood U_x of x and there is $N_x \in \mathbb{Z}^+$ such that for all $n \geq N_x$ we have $x_n \notin U_x$. Indeed, suppose the contrary, there is a point $x \in X$ such that for every neighborhood U_x of x and every integer N there is $n \geq N$ such that x_n in U_x. Let $x_{n_1} \in B(x, 1)$, and for each $k \geq 2$ let n_k be such that $n_k > n_{k-1}$ and $x_{n_k} \in B(x, \frac{1}{k})$. Then $(x_{n_k})_k$ is a subsequence of $(x_n)_n$ convergent to x, a contradiction.

Since the collection $\{U_x \mid x \in X\}$ covers the compact space X it has a finite subcover $\{U_{x_k} \mid 1 \leq k \leq m\}$. Let $N = \max\{N_{x_k} \mid 1 \leq k \leq m\}$. If $n \geq N$ then $x_n \notin U_{x_k}$ for all k, so $x_n \notin X$, a contradiction.

($\Rightarrow$) First we show that for any $\epsilon > 0$ the space X can be covered by finitely many balls of radii ϵ (a property sometimes called **total boundedness**). Suppose the contrary. Take $x_1 \in X$, and inductively take $x_{n+1} \notin \bigcup_{1\leq i\leq n} B(x_i, \epsilon)$.

[1] phủ
[2] phủ con

Since $d(x_m, x_n) \geq \epsilon$ for $m \neq n$, the sequence $(x_n)_{n\geq 1}$ cannot have any convergent subsequence, a contradiction.

Now let O be any open cover of X. By 6.1 there is a corresponding Lebesgue number ϵ such that a ball of radius ϵ is contained in an element of O. The space X is covered by finitely many balls of radii ϵ. The collection of finitely many corresponding elements of O covers X. Thus O has a finite subcover. □

By the above result, in metric spaces compactness can be described by either sequences or open coverings. We inherit all results obtained previously in Analysis about compactness in metric spaces.

Example. If a subspace of a metric space is compact then it is closed and bounded.

A subspace of the Euclidean space $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Compact topological spaces

Generalizing the case of metric spaces in 6.2, for topological spaces the following notion of compactness based on open covers has been found to be appropriate and useful.

Definition. *A topological space is* **compact** *if every open cover has a finite subcover.*

In symbols, a space (X, τ) is compact if

$$\forall I \subset \tau, \bigcup_{O\in I} O = X \implies (\exists J \subset I, |J| < \infty, \bigcup_{O\in J} O = X).$$

Another way to write this statement is

$$\forall I \in \{I \subset \tau \mid \bigcup_{O\in I} O = X\}, \exists J \in \{J \subset I \mid |J| < \infty, \bigcup_{O\in J} O = X\}.$$

From the definition, a space is not compact if and only if there exists an open cover without a finite subcover. In symbols:

$$\exists I \in \{I \subset \tau \mid \bigcup_{O\in I} O = X\}, \{J \subset I \mid |J| < \infty, \bigcup_{O\in J} O = X\} = \emptyset.$$

Remark. A topological space is called **sequentially compact** [1] if every sequence has a convergent subsequence. Theorem 6.2 shows that the notions of compactness and sequentially compactness coincide for metric spaces.

For another approach to a notion of compactness for topological spaces, we may think of replacing sequences by nets and developing a notion of subnets, see page 109.

[1] compắc dãy

Example. Any finite space is compact.

Any space whose topology is finite (that is, the space has finitely many open sets) is compact.

Example. On the Euclidean line $\mathbb{R}$ the collection $\{(-n, n) \mid n \in \mathbb{Z}^+\}$ is an open cover without a finite subcover. Therefore the Euclidean line $\mathbb{R}$ is not compact.

Example. On the Euclidean line $\mathbb{R}$ the the subset $[0, 1]$ has an open cover $\{(-\frac{1}{2}, \frac{2}{3}), (\frac{1}{3}, \frac{3}{2})\}$. If we consider $[0, 1]$ as a space in itself then we have a corresponding open cover $\{[0, \frac{2}{3}), (\frac{1}{3}, 1]\}$.

As an important demonstration we prove:

Proposition. *The Euclidean closed interval* $[0, 1]$ *is compact.*

Proof. Let I be any open cover of $[0, 1]$. Let S be the set of real number x such that the interval $[0, x]$ can be covered by a finite subset of I. The set S is non-empty since $0 \in S$. Let $a = \sup S$. We now show that $a = 1$. Suppose $a < 1$. There is an element U of I which covers a. Since U is open it must contain an interval (b, c) which contains a. From property of supremum the interval (b, c) must contain an element d of S. The interval $[0, d]$ is covered by a finite subset J of I. Then $J \cup \{U\}$ is a finite subset of I which covers $[0, \frac{a+c}{2}]$, a contradiction. □

Theorem (continuous image of compact space is compact). *If X is compact and $f : X \to Y$ is continuous then $f(X)$ is compact.*

Proof. Let I be a cover of $f(X)$ by open sets of $f(X)$. For each $O \in I$ there is U_O open in Y such that $U_O \cap f(X) = O$. Then $\{f^{-1}(U_O) \mid O \in I\}$ is an open cover of X. Since X is compact there is a finite subcover, so there is a finite set $J \subset I$ such that $\{f^{-1}(U_O) \mid O \in J\}$ covers X, $f^{-1}(\bigcup_{O \in J} U_O) = X$, so applying f to both sides we get $\bigcup_{O \in J} U_o \supset f(X)$, hence J is a subcover of I. □

Another consequence is that compactness is preserved under homeomorphism. We say that ***compactness is a topological property***.

Proposition. *Any closed subspace of a compact space is compact.*

Proof. Suppose that X is compact and $A \subset X$ is closed. Let I be a collection of open set of X which by restriction to A gives an open cover of A. By adding the open set $X \setminus A$ to I we get an open cover of X. This open cover has a finite subcover. This subcover of X contains $X \setminus A$, thus omitting this set we get a finite subcover of A from I. □

6.3 Proposition. *In a Hausdorff space a point and a disjoint compact set can be separated by open sets.*

Proof. Let A be a compact set in a Hausdorff space X. Let $x \in X \setminus A$. For each $a \in A$ there are disjoint open sets U_a containing x and V_a containing a. The collection $\{V_a \mid a \in A\}$ covers A, so there is a finite subcover $\{V_{a_i} \mid 1 \leq i \leq n\}$. Let $U = \bigcap_{i=1}^n U_{a_i}$ and $V = \bigcup_{i=1}^n V_{a_i}$. Then U is an open neighborhood of x disjoint from V, a neighborhood of A. □

We deduce immediately from 6.3:

6.4 Corollary. *Any compact subspace of a Hausdorff space is closed.*

Example. Any subspace of $\mathbb{R}$ with the finite complement topology, closed or not, is compact, see 6.16. This space is not Hausdorff, see 5.7. Thus ***in general topological spaces it is not true that being compact implies being closed***.

We easily derive a familiar theorem that a real-valued function on a compact space achieves extrema:

Theorem (Extreme value theorem). *If X is a compact space and $f : X \to \mathbb{R}$ where $\mathbb{R}$ has the Euclidean topology is continuous then f has a maximum value and a minimum value.*

Proof. Since X is compact and f is continuous, $f(X)$ is compact. Under the Euclidean topology of $\mathbb{R}$, which is the topology generated by the Euclidean metric, a compact subset must be closed and bounded. Boundedness of $f(X)$ implies $\inf f(X)$ and $\sup f(X)$ exist, and closeness of $f(X)$ implies $\inf f(X)$ and $\sup f(X)$, which are contact points of $f(X)$, belong to $f(X)$, so they are $\min f(X)$ and $\max f(X)$. □

Characterization of compact spaces in terms of closed subsets

In the definition of compact spaces by writing open sets as complements of closed sets, we get a dual statement: A space is compact if for every collection of closed subsets whose intersection is empty there is a a finite subcollection whose intersection is empty. A collection of subsets of a set has the **finite intersection property** [1] if the intersection of every finite subcollection is non-empty. We get:

6.5 Proposition. *A space is compact if and only if every collection of closed subsets with the finite intersection property has non-empty intersection.*

Proof. Let F be a collection of subsets of X, then $\bigcap_{C \in F} C = \emptyset$ if and only if $\bigcup_{C \in F}(X \setminus C) = X$. The collection F does not have the finite intersection property if and only if there is a finite set $I \subset F$ such that $\bigcap_{C \in I} C = \emptyset$, which

[1] tính giao hữu hạn

is equivalent to that $\bigcup_{C \in I}(X \setminus C) = X$. Hence the existence of a collection of closed subsets with finite intersection property and with empty intersection is equivalent to the existence of a collection of open subsets which covers X with no finite subcover. □

Compactification

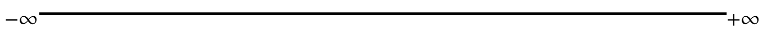
A **compactification** [1] of a space X is a compact space Y such that X is homeomorphic to a dense subspace of Y.

Example. The Euclidean interval $[0, 1]$ is a compactification of the Euclidean interval $[0, 1)$.

The Euclidean interval $[0, 1]$ is also a compactification of the Euclidean interval $(0, 1)$.

6.6 Example. Since the Euclidean line $\mathbb{R}$ is homeomorphic to the interval $(-1, 1)$, it can be compactified by adding two elements $+\infty$ and $-\infty$ and equip a topology on $\mathbb{R} \cup \{\pm\infty\}$ such that it is homeomorphic to the interval $[-1, 1]$, a compact space. Details can be worked out in

This set is sometimes called the extended set of real numbers, is used since highschool mathematics, and is often drawn as an extended straight line.



In Calculus we write such formulas as

$$\lim_{x \to +\infty} x^2 = +\infty.$$

This means, by definition,

$$\forall \epsilon > 0, \exists \delta > 0, \forall x > \delta, x^2 > \epsilon.$$

Writing in terms of open sets,

$$\forall (\epsilon, +\infty), \exists (\delta, +\infty), \forall x \in (\delta, +\infty), x^2 \in (\epsilon, +\infty).$$

Thus it can be understood that intervals of the form $(\epsilon, +\infty)$ are open neighborhoods of $+\infty$.

Example. A compactification of the Euclidean interval $(0, 1)$ is the circle S^1, by adding just one point, see Fig. 6.7. We can also say that the circle is a compactification of the Euclidean line by adding one point. In other words, $\mathbb{R} \cup \{\infty\}$ has a topology such that it contains $\mathbb{R}$ as a subspace and is homeomorphic to S^1.

[1] compắc hóa

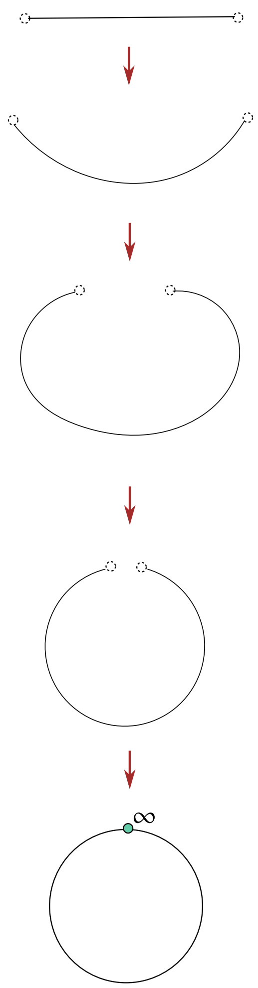


Figure 6.7: A one-point compactification of the Euclidean interval $(0, 1)$.

Example. The Euclidean plane $\mathbb{R}^2$ is homeomorphic to the two-dimensional sphere S^2 minus one point. Thus the two-dimensional sphere S^2 is a compactification of $\mathbb{R}^2$, by adding just one point, which can be considered as a point at infinity ∞, to $\mathbb{R}^2$.

We now investigate whether it is possible to do this in general. Figure 6.7 is a template for this construction.

Let X be a non-empty space. Since the collection $\mathcal{P}(X)$ of all subsets of X cannot be contained in X (see 1.9) there is an element of $\mathcal{P}(X)$ which is not in X. Let us denote that element by ∞, and let $X^\infty = X \cup \{\infty\}$.

Let us see what a topology on X^∞ should be in order for X^∞ to contain X as a subspace and to be compact. That X^∞ contains X as a subspace means the subspace topology of X with respect to X^∞ is the original topology of X. If an open subset U of X^∞ does not contain ∞ then U is contained in X, therefore U is an open subset of X in the subspace topology of X. If U contains ∞ then the complement $X^\infty \setminus U$ must be a closed subset of X^∞, hence $X^\infty \setminus U$ is compact in the subspace topology from X^∞, furthermore $X^\infty \setminus U$ is contained in X and is therefore a closed subset of X in the subspace topology of X. In summary, an open subset of X^∞ must be either an open subset of X or the complement of a closed compact subset of X.

Example. Give X^∞ the topology $\tau_X \cup \{X\}$. Then X^∞ is compact.

The following is an important compactification: [1]

Theorem. *The collection consisting of all open subsets of X and all complements in X^∞ of closed compact subsets of X is the finest topology on X^∞ such that X^∞ is compact and contains X as a subspace. If X is not compact then X is dense in X^∞, and X^∞ is called the* **Alexandroff compactification** *of X.*

Proof. We go through several steps.

(a) We check that we truely have a topology on X^∞.

- On unions: Let I be a collection of closed compact sets in X. Then $\bigcup_{C\in I}(X^\infty \setminus C) = X^\infty \setminus \bigcap_{C\in I} C$, where $\bigcap_{C\in I} C$ is closed compact. If U is open in X and C is closed compact in X then $U \cup (X^\infty \setminus C) = X^\infty \setminus (C \setminus U)$, where $C \setminus U$ is closed compact in X.
- On finite intersections: If U is open in X and C is closed compact in X then $U \cap (X^\infty \setminus C) = U \cap (X \setminus C)$ is open in X. If C_1 and C_2 are closed compact in X then $(X^\infty \setminus C_1) \cap (X^\infty \setminus C_2) = X^\infty \setminus (C_1 \cup C_2)$, where $C_1 \cup C_2$ is closed compact in X.

(b) We check that with this topology X is a subspace of X^∞. This means the original topology of X coincides with the subspace topology which X receives from X^∞.

- Open subsets in the original topology of X are open subsets of X^∞, hence are open in the subspace topology.
- Consider restrictions of open subsets of X^∞ to X. If an open subset U of X^∞ does not contain ∞ then U is an open subset in the original topology of X, while the restriction of U to X is U, giving that U is an open subset of X in the subspace topology. If U contains ∞ then $U = X^\infty \setminus C$ where C is closed compact in the original topology of X, then $U \cap X = (X^\infty \setminus C) \cap X = X \setminus C$ is open in the original topology of X.

(c) We check that X^∞ is compact. Let F be an open cover of X^∞. Then an element $O \in F$ will cover ∞. The complement of O in X^∞ is a compact set C in X. Then $F \setminus \{O\}$ is an open cover of C. From this cover there is a finite cover. This finite cover together with O is a finite cover of X^∞.

[1] Proved in the early 1920s by Pavel Sergeyevich Alexandrov. Alexandroff is another spelling of this name.

(d) The argument preceding this theorem confirms that this topology contains any topology such that X^∞ is compact and contains X as a subspace.

(e) Since X is not compact and X^∞ is compact, X cannot be closed in X^∞, therefore the closure of X in X^∞ is X^∞.

□

Sometimes the abstract Alexandroff compactification can be identified, that is, it is found to be homeomorphic to a more familiar space, by using the following result:

6.8 Proposition. *Suppose that a space $X^\infty = X \cup \{\infty\}$ is compact and contains X as a subspace, and is a Hausdorff space. Then X^∞ is the Alexandroff compactification of X. If a space Y is homeomorphic to X then the Alexandroff compactification of Y is homeomorphic to X^∞.*

Example. The Euclidean line $\mathbb{R}$ is homeomorphic to the circle S^1 minus a point. The circle is a compact Hausdorff space. By 6.8, the Alexandroff compactification of the Euclidean line is homeomorphic to the circle.

Example. Similarly, the Euclidean plane $\mathbb{R}^2$ is homeomorphic to the sphere S^2 minus a point. Since S^2 is a compact Hausdorff space, the Alexandroff compactification of the Euclidean plane $\mathbb{R}^2$ is homeomorphic to the sphere S^2. When $\mathbb{R}^2$ is identified with the complex plane $\mathbb{C}$, the sphere S^2 is often called the **Riemann sphere**.

Proof. We check that the topology of X^∞ is the topology of the Alexandroff compactification.

Let U be open in the original topology of X. Then U is also open in the subspace topology of X relative to X^∞ so there is an open subset V in X^∞ such that $V \cap X = U$. That X^∞ is a Hausdorff space implies $\{\infty\}$ is closed in X^∞, hence $X = X^\infty \setminus \{\infty\}$ is open in X^∞. Thus U is open in X^∞.

Let C be closed compact in the original topology of X. Then C is a compact subspace of X^∞ which is a Hausdorff space, hence C is closed in X^∞, by 6.4. This implies $X^\infty \setminus C$ is open in X^∞.

We find that the topology of X^∞ contains the topology of the Alexandroff compactification. Since the topology of the Alexandroff compactification is the finest topology such that X^∞ is compact and contains X as a subspace, the topology of X^∞ is exactly the topology of the Alexandroff compactification.

Let $h : Y \to X$ be a homeomorphism. Let $Y \cup \{b\}$ be the Alexandroff compactifications of Y. Let $\tilde{h} : Y \cup \{b\} \to X^\infty$ be defined by $\tilde{h}(y) = h(y)$ if $y \neq b$ and $\tilde{h}(b) = \infty$. This is a bijection. We check that it is a homeomorphism.

We check that $\tilde{h}$ is continuous. Let U be an open subset of X^∞. If U does not contain ∞ then U is open in X, so $\tilde{h}^{-1}(U) = h^{-1}(U)$ is open in Y, and so $\tilde{h}^{-1}(U)$ is open in $Y \cup \{b\}$. If U contains ∞ then $X^\infty \setminus U$ is closed compact in X, so $h^{-1}(X^\infty \setminus U)$ is closed compact in X, hence $\tilde{h}^{-1}(U) = (Y \cup \{b\}) \setminus h^{-1}(X^\infty \setminus U)$ is open in $Y \cup \{b\}$.

That the inverse map is continuous is similar, or we can use 6.14 instead. □

* Locally compact spaces

A topological space is **locally compact** if every point has a compact neighborhood.

Example. The Euclidean space $\mathbb{R}^n$ is locally compact.

Proposition. *The Alexandroff compactification of a space X is a Hausdorff space if and only if X is a locally compact Hausdorff space.*

Proof. Suppose that X is locally compact Hausdorff space. We only need to check that in the Alexandroff compactification X^∞ the points ∞ and $x \in X$ can be separated by open sets. Since X is locally compact there is a compact set C containing an open neighborhood O of x. Since X is a Hausdorff space, C is closed in X. Then from the definition of the Alexandroff compactification, $X^\infty \setminus C$ is open in X^∞. So O and $X^\infty \setminus C$ separate x and ∞.

If the Alexandroff compactification X^∞ is a Hausdorff space then X as a subspace is also a Hausdorff space (5.23). Given a point $x \in X$, since the points x and ∞ can be separated by open sets of X^∞, there is an open set $O \subset X$ containing x and a closed compact set $C \subset X$ such that $O \cap (X^\infty \setminus C) = \emptyset$. Then $x \in O \subset C$, and C is a compact neighborhood of X. Thus X is locally compact. □

Problems

6.9. Show that any discrete compact topological space is finite.

6.10. Show that in a topological space any finite unions of compact subspaces is compact.

6.11. Show that in a Hausdorff space any intersection of compact subspaces is compact.

6.12 (extension of Cantor lemma in Calculus). Show that if X is compact and $X \supset A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ is a descending sequence of closed, non-empty sets, then $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$.

6.13 (continuous functions on compact metric spaces are uniformly continuous). A function f from a metric space to a metric space is **uniformly continuous** if for any $\epsilon > 0$, there is $\delta > 0$ such that if $d(x, y) < \delta$ then $d(f(x), f(y)) < \epsilon$. Prove that a continuous function from a compact metric space to a metric space is uniformly continuous, a result often encountered in Analysis, by using open covering compactness.

6.14. ✓ Show that if X is compact, Y is Hausdorff, $f : X \to Y$ is bijective and continuous, then f is a homeomorphism.

6.15. Show that in a compact space any infinite set has a limit point.

6.16. Show that any subspace of $\mathbb{R}$ with the finite complement topology is compact. Deduce that *there is a subspace that is compact but is not closed*.

6.17. Show that in a regular space a closed set and a disjoint compact set can be separated by open sets.

6.18. ✓ Show that in a Hausdorff space two disjoint compact sets can be separated by open sets.

6.19. ✓ Show that any compact Hausdorff space is normal.

6.20. Prove the existence of Lebesgue number for compact metric spaces 6.1 by using open covering compactness.

6.21 (another proof of the existence of Lebesgue number). Suppose $(U_\alpha)_{\alpha\in A}$ is an open covering of a sequentially compact metric space X.

(a) For $x \in X$ let $f(x) = \sup\{r \in \mathbb{R} \mid \exists \alpha \in A, B(x,r) \subset U_\alpha\}$. Show that f is well-defined, positive on X.

(b) Check that for any x and y we have $f(y) \leq f(x) + d(x,y)$. Deduce that f is continuous.

(c) Take $0 < \epsilon < \min_{x\in X} f(x)$.

This argument is given in [BBI01, p. 15].

6.22. Check that the Topologist's sine curve $\{(x, \sin\frac{1}{x}) \mid 0 < x \leq 1\} \cup \{(0,y) \mid -1 \leq y \leq 1\}$ (see 4.9) is a compactification of the Euclidean interval $(0,1)$.

6.23. Find the Alexandroff compactification of $\{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$ under the Euclidean topology.

6.24. Find the Alexandroff compactification of $\mathbb{Z}^+$ under the Euclidean topology. This means to find a familiar space which is homeomorphic to the abstract Alexandroff compactification of $\mathbb{Z}^+$.

6.25. Find the Alexandroff compactification of the Euclidean space $\mathbb{R}^n$.

6.26. Find the Alexandroff compactification of the Euclidean open ball $B(0,1)$?

6.27 (extension of functions to compactifications). Consider $\mathbb{R}$ with the Euclidean topology. Find a necessary and sufficient condition for a continuous function from $\mathbb{R}$ to $\mathbb{R}$ to have an extension to a continuous function from the Alexandroff compactification $\mathbb{R} \cup \{\infty\}$ to $\mathbb{R}$.

6.28 (extended set of real numbers). Continue Example 6.6.

(a) Define a topology on $\overline{\mathbb{R}} = \mathbb{R} \cup \pm\infty$ such that the following map (see 3.4)

$$\begin{aligned} h : \mathbb{R} &\mapsto (-1,1) \\ x &\mapsto h(x) = \frac{x}{\sqrt{1+x^2}} \end{aligned}$$

induces a homemorphism from $\overline{\mathbb{R}}$ onto $[-1,1]$.

(b) Let $f(x) = e^{-x^2}$, $x \in \mathbb{R}$. Find an extension of f to a continuous map $\overline{f} : \overline{\mathbb{R}} \to \mathbb{R}$.

(c) Generalize, give a sufficient condition for a continuous map $f : \mathbb{R} \to \mathbb{R}$ to have an extension to a continuous map $\overline{f} : \overline{\mathbb{R}} \to \mathbb{R}$.

6.29. Similarly to Example 6.6, since the Euclidean $\mathbb{R}^n$ is homeomorphic to the ball $B(0, 1)$ by 3.4, find a compactification of $\mathbb{R}^n$ which is homeomorphic to the closed ball $B'(0, 1)$.

6.30. Show that $\mathbb{Q}$ is not locally compact under the Euclidean topology of $\mathbb{R}$.

6.31. * We could have noticed that the notion of local compactness as we have defined is not apparently a local property. For a property to be local, every neighborhood of any point must contain a neighborhood of that point with the given property (as in the cases of local connectedness and local path-connectedness). Show that for Hausdorff spaces local compactness is indeed a local property, i.e., every neighborhood of any point contains a compact neighborhood of that point.

6.32. Any locally compact Hausdorff space is a regular space.

6.33. In a locally compact Hausdorff space, if K is compact, U is open, and $K \subset U$, then there is an open set V such that $\overline{V}$ is compact and $K \subset V \subset \overline{V} \subset U$.

6.34. * A space is a locally compact Hausdorff space if and only if it is homeomorphic to an open subspace of a compact Hausdorff space.

6.35. * Let X be a compact Hausdorff space. Let $X_1 \supset X_2 \supset \cdots$ be a nested sequence of closed connected subsets of X. Show that $\bigcap_{i=1}^{\infty} X_i$ is connected.

7 Product of spaces

Finite products of metric spaces

Example (Euclidean topology). Let $\mathbb{R}$ have Euclidean topology, generated by open intervals. Consider the set $\mathbb{R}^n$ as $\mathbb{R}^n = \underbrace{\mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}}_{n \text{ copies of } \mathbb{R}}$. Recall that the open rectangles form a basis for the Euclidean topology of $\mathbb{R}^n$, see 2.2 and 2.22, any non-empty open set in the Euclidean topology of $\mathbb{R}^n$ is a union of products of open intervals of $\mathbb{R}$.

In Analysis the following metric on finite product of metric spaces is often used. Let (X_i, d_i), $1 \le i \le n$ be metric spaces. Let $X = \prod_{i=1}^n X_i$. For $x = (x_1, x_2, \dots, x_n) \in X$ and $y = (y_1, y_2, \dots, y_n) \in X$, define

$$\delta(x,y) = \left(\sum_{i=1}^n d_i(x_i,y_i)^2\right)^{1/2}.$$

We can check that δ is a metric on X (Problem 7.10).

Given a ball $B_\delta(x,r)$ in the product space, we can find $\prod_{i=1}^n B_{d_i}\left(x_i, \frac{r}{\sqrt{n}}\right)$ as an open set contained in $B_\delta(x,r)$. On the other hand any set in the product space of the form $\prod_{i=1}^n B_{d_i}(x_i, r_i)$ contains a ball $B_\delta(x,r)$ with $r = \min\{r_i \mid 1 \le i \le n\}$. By 2.22 the topology generated by the metric δ is the same as the topology generated by the collections of products of balls in each component spaces.

Recall, and we can check again quickly, that this topology gives an important property, that the projection maps $p_j : \prod_{i=1}^n X_i \to X_j$, $p_j((x_i)_{1\le i\le n}) = x_j$ are continuous.

Another important property is that a product of compact metric spaces is compact (Problem 7.27). This result leads to the characterization of compact subsets of Euclidean spaces as closed bounded subsets (Problem 7.26).

Finite products of topological spaces

From the case of products of metric spaces, we propose a topology on product of topological spaces.

Definition. *Let X and Y be two topological spaces. The* **product topology** *on the set $X \times Y$ is the topology generated by the collection of sets of the form $U \times V$ where U is an open set of X and V is an open set of Y.*

Similarly, the product topology on a finite product of topological spaces $\prod_{i=1}^n (X_i, \tau_i)$ is defined to be the topology generated by the collection $F = \{\prod_{i=1}^n U_i \mid U_i \in \tau_i\}$.

Since the intersection of two members of F is also a member of F, the collection F is a basis for the product topology. Thus ***open sets in the product***

topology are unions of products of open sets.

Proposition. *If each b_i is a basis for X_i then $\{\prod_{i=1}^n U_i \mid U_i \in b_i\}$ is a basis for the product topology on $\prod_{i=1}^n X_i$.*

Proof. Consider an element in the above basis of the product topology, which is of the form $\prod_{i=1}^n V_i$ where $V_i \in \tau_i$. Each V_i can be written $V_i = \bigcup_{i_j \in I_i} U_{i_j}$, where $U_{i_j} \in b_i$. Then

$$\prod_{i=1}^n V_i = \bigcup_{i_j \in I_i, 1 \le i \le n} \prod_{i=1}^n U_{i_j}.$$

□

Example (Euclidean topology). Recall that $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ copies of } \mathbb{R}} = \prod_{i=1}^n \mathbb{R}$. Let $\mathbb{R}$ have Euclidean topology, generated by open intervals. An open set in the product topology of $\mathbb{R}^n$ is a union of products of open intervals. Since a product of open intervals is an open rectangle, and the open rectangles form a basis for the Euclidean topology, see 2.2 and 2.22, ***the product topology on $\mathbb{R}^n$ is exactly the Euclidean topology***.

7.1 Example (finite products of metric spaces). In Analysis the following metric on finite product of metric spaces is often used. Let (X_i, d_i), $1 \le i \le n$ be metric spaces. Let $X = \prod_{i=1}^n X_i$. For $x = (x_1, x_2, \dots, x_n) \in X$ and $y = (y_1, y_2, \dots, y_n) \in X$, define

$$\delta_2(x, y) = \left(\sum_{i=1}^n d_i(x_i, y_i)^2 \right)^{1/2}.$$

We can check that δ_2 is a metric on X (Problem 7.10).

Given a ball $B(x, r)$ in (X, δ_2), we can find $\prod_{i=1}^n B\left(x_i, \frac{r}{\sqrt{n}}\right)$ as an open set in the product topology contained in $B(x, r)$. On the other hand an open set in the product topology of the form $\prod_{i=1}^n B(x_i, r_i)$ contains a ball $B(x, r)$ with $r = \min\{r_i \mid 1 \le i \le n\}$. By 2.22 the topology generated by the metric δ_2 is exactly the product topology. Thus finite products of metric spaces are special cases of finite products of topological spaces.

Remark. Note a common error: it is ***not true that an arbitrary open set in the product topology is a product of open sets***.

In the Euclidean plane, an open disk cannot be a product of two sets, the union of two rectangles may not be a product of two sets, see Fig. 7.2.

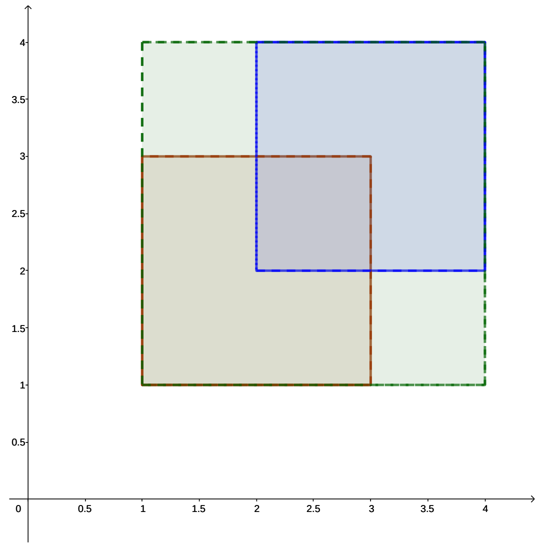


Figure 7.2: In the Euclidean plane the union of the two rectangles $(1,3)\times(1,3)$ and $(2,4)\times(2,4)$ is an open set S. If S is a product $U\times V$ then the set of first coordinates of S must be U, thus $U=(1,4)$, and similarly $V=(1,4)$, but clearly S is not $(1,4)\times(1,4)$, thus S is not a product of two sets.

Arbitrary products of spaces

Definition. *Let $((X_i,\tau_i))_{i\in I}$ be an indexed family of topological spaces. The* **product topology** *on the set $\prod_{i\in I} X_i$ is the topology generated by the collection F consisting of all sets of the form $\prod_{i\in I} U_i$, where $U_i\in\tau_i$ and $U_i=X_i$ for all except finitely many $i\in I$. In symbols:*

$$F=\left\{\prod_{i\in I} U_i \mid U_i\in\tau_i, \exists J\subset I, |J|<\infty, \forall i\in I\setminus J, U_i=X_i\right\}.$$

The collection F is a basis of the product topology since it is closed under finite intersections. Thus ***any open set is a union of open sets $\prod_{i\in I} U_i$ where only finitely many of U_i are not the whole space X_i***.

The following subcollection of F,

$$G=\left\{\prod_{i\in I} U_i \mid U_i\in\tau_i, \exists j\in I, \forall i\in I\setminus\{j\}, U_i=X_i\right\}, \tag{7.3}$$

that is, the collection of all sets of the form $\prod_{i\in I} U_i$, where $U_i\in\tau_i$ and $U_i=X_i$ for all except one $i\in I$, is a subbasis for the product topology, since every element of F is the intersection of a finite subcollection of G.

Example. The set of all sequences of real numbers $\{(x_1,x_2,\dots,x_n,\dots)\mid n\in\mathbb{Z}^+, x_n\in\mathbb{R}\}$, often considered in Analysis, is exactly the product of an infinitely countable collection of the set of real numbers, $\prod_{n\in\mathbb{Z}^+}\mathbb{R}$, which may be denoted by $\mathbb{R}^{\mathbb{Z}^+}$ for similarity in notation to $\mathbb{R}^n$. It can be equipped with the product topology.

Recall that an element of the set $\prod_{i\in I} X_i$ is written $(x_i)_{i\in I}$ (see Section 1 on page 6). For $j \in I$ the **projection to the j-coordinate** is defined by $p_j : \prod_{i\in I} X_i \to X_j$, $p_j((x_i)_{i\in I}) = x_j$.

The definition of the product topology is explained in the following:

7.4 Theorem (product topology is the topology such that projections are continuous). *The product topology is the coarsest topology on $\prod_{i\in I} X_i$ such that all the projection maps p_i are continuous. In other words, the product topology is the topology generated by the projection maps.*

Proof. Notice that if $O_j \subset X_j$ then $p_j^{-1}(O_j) = \prod_{i\in I} U_i$ with $U_i = X_i$ for all i except j, and $U_j = O_j$. The topology generated by all the maps p_i is the topology generated by all sets of the form $p_i^{-1}(O_i)$ with $O_i \in \tau_i$, which is the collection G in Equation (7.3). □

Fore more discussions on topologies generated by maps, see Section 9.1 on page 96.

7.5 Theorem (map to a product space is continuous if and only if each component map is continuous). *A map $f : Y \to \prod_{i\in I} X_i$ is continuous if and only if each component $f_i = p_i \circ f$ is continuous.*

Proof. If $f : Y \to \prod_{i\in I} X_i$ is continuous then $p_i \circ f$ is continuous because p_i is continuous.

On the other hand, let us assume that every f_i is continuous. Let $U = \prod_{i\in I} U_i$ where $U_i \in \tau_i$ and $\exists j \in I$ such that $U_i = X_i, \forall i \neq j$ be an element of the subbasis of $\prod_{i\in I} X_i$. Then

$$\begin{aligned} f^{-1}(U) &= \{y \in Y \mid f(y) \in U\} = \{y \in Y \mid \forall i \in I, f(y)_i \in U_i\} \\ &= \{y \in Y \mid f(y)_j \in U_j\} = \{y \in Y \mid f_j(y) \in U_j\} = f_j^{-1}(U_j). \end{aligned}$$

is an open set since f_j is continuous. So f is continuous. □

7.6 Theorem (convergence in product topology is coordinate-wise convergence). *A net $\left(\left(x_i^j\right)_{i\in I}\right)_{j\in J}$ in $\prod_{i\in I} X_i$ (a map $n : J \to \prod_{i\in I} X_i$) is convergent if and only if each component net $\left(x_i^j\right)_{j\in J}$ (the projection $p_i \circ n$) is convergent.*

Proof. ($\Rightarrow$) This is because the projection maps are continuous, and by 5.5.

($\Leftarrow$) Suppose that each $p_i \circ n$ is convergent to a_i, we show that n is convergent to $a = (a_i)_{i\in I}$.

A neighborhood of a contains an open set of the form $U = \prod_{i\in I} O_i$ with O_i are open sets of X_i and $O_i = X_i$ except for $i \in K$, where K is a finite subset of I.

For each $i \in K$, $p_i \circ n$ is convergent to a_i, therefore there exists an index $j_i \in J$ such that for $j \geq j_i$ we have $p_i(n(j)) \in O_i$. Since J is directed, take an index j_0 such that $j_0 \geq j_i$ for all $i \in K$. Then for $j \geq j_0$ we have $n(j) \in U$. □

Tikhonov theorem

The general result for arbitrary product of compact topological spaces is called Tikhonov theorem [1]:

Theorem (Tikhonov theorem – A product of compact spaces is compact). *If X_i is compact for all $i \in I$ then $\prod_{i\in I} X_i$ is compact.*

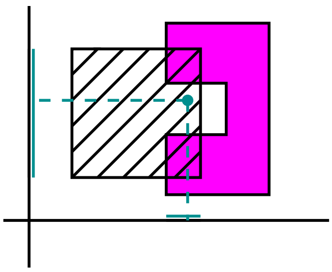
Example. With the Euclidean topology on $[0,1]$, the space $[0,1]^{\mathbb{Z}^+} = \prod_{i\in\mathbb{Z}^+}[0,1]$ is called the **Hilbert cube**. By Tikhonov theorem the Hilbert cube is compact.

In the case of finite product of topological spaces the result can be proved more easily than in the general case (Problem 7.22).

Proof of Tikhonov theorem. Let X_i be compact for all $i \in I$. We will show that $X = \prod_{i\in I} X_i$ is compact by showing that if a collection of closed subsets of X has the finite intersection property then it has non-empty intersection (see 6.5). (A proof based on open covers is also possible, see [Kel55, p. 143].)

Let F be a collection of closed subsets of X that has the finite intersection property. We will show that $\bigcap_{A\in F} A \neq \emptyset$.

Have a look at the following argument, which suggests that proving the Tikhonov theorem might not be easy. If we take the closures of the projections of the collection F to the i-coordinate then we get a collection $\{\overline{p_i(A)}, A \in F\}$ of closed subsets of X_i having the finite intersection property. Since X_i is compact, this collection has non-empty intersection. From this it is tempting to conclude that F must have non-empty intersection itself. But that is not true, see the following figure.



We shall overcome this difficulty by first enlarging the collection F.

(a) We show that there is a maximal collection $\tilde{F}$ of subsets of X such that $\tilde{F}$ contains F and still has the finite intersection property. We will use Zorn lemma for this purpose. (This is a routine step, it might be easier to carry it out than to read.)

[1] Proved by Andrei Nicolaievich Tikhonov around 1926. The product topology was defined by him. His name is also spelled as Tychonoff.

Let K be the collection of collections G of subsets of X such that G contains F and G has the finite intersection property. On K we define an order by the usual set inclusion.

Now suppose that L is a totally ordered subcollection of K. Let $H = \bigcup_{G\in L} G$. We will show that $H \in K$, therefore H is an upper bound of L.

First H contains F. We need to show that H has the finite intersection property. Suppose that $H_i \in H$, $1 \leq i \leq n$. Then $H_i \in G_i$ for some $G_i \in L$. Since L is totally ordered, there is an i_0, $1 \leq i_0 \leq n$ such that G_{i_0} contains all G_i, $1 \leq i \leq n$. Then $H_i \in G_{i_0}$ for all $1 \leq i \leq n$, and since G_{i_0} has the finite intersection property, we have $\bigcap_{i=1}^{n} H_i \neq \emptyset$.

(b) Since $\tilde{F}$ is maximal, it is closed under finite intersection. Moreover if a subset of X has non-empty intersection with every element of $\tilde{F}$ then it belongs to $\tilde{F}$.

(c) Since $\tilde{F}$ has the finite intersection property, for each $i \in I$ the collection $\{p_i(A) \mid A \in \tilde{F}\}$ also has the finite intersection property, and so does the collection $\{\overline{p_i(A)} \mid A \in \tilde{F}\}$. Since X_i is compact, $\bigcap_{A\in\tilde{F}} \overline{p_i(A)}$ is non-empty.

(d) Let $x_i \in \bigcap_{A\in\tilde{F}} \overline{p_i(A)}$ and let $x = (x_i)_{i\in I} \in \prod_{i\in I}[\bigcap_{A\in\tilde{F}} \overline{p_i(A)}]$. We will show that $x \in \overline{A}$ for all $A \in \tilde{F}$, in particular $x \in A$ for all $A \in F$.

We need to show that any neighborhood of x has non-empty intersection with every $A \in \tilde{F}$. It is sufficient to prove this for neighborhoods of x belonging to the basis of X, namely finite intersections of sets of the form $p_i^{-1}(O_i)$ where O_i is an open neighborhood of $x_i = p_i(x)$. For any $A \in \tilde{F}$, since $x_i \in \overline{p_i(A)}$ we have $O_i \cap p_i(A) \neq \emptyset$. Therefore $p_i^{-1}(O_i) \cap A \neq \emptyset$. By the maximality of $\tilde{F}$ under finite intersection property, we have $p_i^{-1}(O_i) \in \tilde{F}$, and the desired result follows.

□

Applications of Tikhonov theorem include the Stone-Cech compactification 9.10 and the Banach–Alaoglu theorem in Functional Analysis [Bre11].

Problems

7.7. Note that, as sets:

(a) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

(b) $(A \times B) \cup (C \times D) \subsetneq (A \cup C) \times (B \cup D) = (A \times B) \cup (A \times D) \cup (C \times B) \cup (C \times D)$.

7.8. Check that in topological sense (i.e. up to homeomorphisms):

(a) $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

(b) $(X \times Y) \times Z = X \times Y \times Z$.

(c) $(X \times Y) \times Z = X \times (Y \times Z)$.

7.9. How are $X \times Y$ and $Y \times X$ related?

7.10. Continue 7.1, let (X_i, d_i), $1 \le i \le n$ be metric spaces. Let $X = \prod_{i=1}^n X_i$. For $x = (x_1, x_2, \dots, x_n) \in X$ and $y = (y_1, y_2, \dots, y_n) \in X$, define

$$\delta_1(x, y) = \sum_{i=1}^n d_i(x_i, y_i)$$

$$\delta_2(x, y) = \left(\sum_{i=1}^n d_i(x_i, y_i)^2\right)^{1/2}$$

$$\delta_\infty(x, y) = \max\{d_i(x_i, y_i) \mid 1 \le i \le n\}.$$

Check that these are metrics on X. Check that these metrics are equivalent, hence generate the same topology – the product topology.

7.11. If for each $i \in I$ the space X_i is homeomorphic to the space Y_i then $\prod_{i \in I} X_i$ is homeomorphic to $\prod_{i \in I} Y_i$.

7.12. ✓ Show that ***the graph of a continuous function is homeomorphic to its domain***. Namely, let $f : X \to Y$ be continuous, and let $\text{graph}(f) = \{(x, f(x)) \mid x \in X\} \subset X \times Y$, then $\text{graph}(f)$ is homeomorphic to X.

7.13. Is a product of open sets open? Is a product of closed sets closed?

7.14. Is a projection of an open set open? In other words, is the projection map p_i an open map, bringing open sets to open sets? Is a projection of a closed set closed?

7.15. ✓ Show that a space X is Hausdorff if and only if the diagonal $\Delta = \{(x, x) \in X \times X\}$ is closed in $X \times X$, by:

(a) using nets,

(b) not using nets.

7.16. Show that if Y is Hausdorff and $f : X \to Y$ is continuous then the graph of f, that is, the set $\{(x, f(x)) \mid x \in X\}$, is closed in $X \times Y$, by:

(a) using nets,

(b) not using nets.

7.17. Let X be Hausdorff and Y be compact. Suppose the graph of the map $f : X \to Y$ is closed in $X \times Y$. Prove that f is continuous.

7.18. ✓ Fix a point $O = (O_i) \in \prod_{i \in I} X_i$. Define the inclusion map $f : X_i \to \prod_{i \in I} X_i$ by

$$x \mapsto f(x) \text{ with } f(x)_j = \begin{cases} O_j & \text{if } j \neq i \\ x & \text{if } j = i \end{cases}.$$

Show that f is a homeomorphism onto its image $\tilde{X}_i$ (an embedding of X_i). Thus $\tilde{X}_i$ is a copy of X_i in $\prod_{i \in I} X_i$. The spaces $\tilde{X}_i$ have O as the common point. This is analogous to the coordinate system Oxy on $\mathbb{R}^2$.

7.19. ✓ Show that

(a) If each X_i, $i \in I$ is a Hausdorff space then $\prod_{i\in I} X_i$ is a Hausdorff space.

(b) If $\prod_{i\in I} X_i$ is a Hausdorff space then each X_i is a Hausdorff space.

7.20. Show that

(a) If $\prod_{i\in I} X_i$ is path-connected then each X_i is path-connected.

(b) If each X_i, $i \in I$ is path-connected then $\prod_{i\in I} X_i$ is path-connected.

7.21. ✓ Show that

(a) If $\prod_{i\in I} X_i$ is connected then each X_i is connected.

(b) If X and Y are connected then $X \times Y$ is connected.

(c) * If each X_i, $i \in I$ is connected then $\prod_{i\in I} X_i$ is connected.

7.22. Show that

(a) If $\prod_{i\in I} X_i$ is compact then each X_i is compact.

(b) * If X and Y are compact then $X \times Y$ is compact (of course without using the Tikhonov theorem).

7.23. Show that the sphere S^2 with the North Pole and the South Pole removed is homeomorphic to the infinite cylinder $S^1 \times \mathbb{R}$.

7.24 (disjoint union). ✓ Let A_1 and A_2 be topological spaces. On the set $(A_1 \times \{1\}) \cup (A_2 \times \{2\})$ consider the topology generated by subsets of the form $U \times \{1\}$ and $V \times \{2\}$ where U is open in A_1 and V is open in A_2. Show that $A_1 \times \{1\}$ is homeomorphic to A_1, while $A_2 \times \{2\}$ is homeomorphic to A_2. Notice that $(A_1 \times \{1\}) \cap (A_2 \times \{2\}) = \emptyset$. The space $(A_1 \times \{1\}) \cup (A_2 \times \{2\})$ is called the **disjoint union** [1] of A_1 and A_2, denoted by $A_1 \sqcup A_2$. We can use this construction when for example we want to consider a space consisting of two disjoint circles.

More generally, given a collection of spaces $((X_i, \tau_i))_{i\in I}$ the disjoint union of this collection is the space $\bigcup_{i\in I}(X_i \times \{i\})$ with the topology $\bigcup_{i\in I} \tau_i \times \{i\}$. Check that this is indeed a topology.

7.25. Let X be a normed space over a field $\mathbb{F}$ which is $\mathbb{R}$ or $\mathbb{C}$. Check that the addition $(x, y) \mapsto x + y$ is a continuous map from the product space $X \times X$ to X, while the scalar multiplication $(c, x) \mapsto c \cdot x$ is a continuous map from $\mathbb{F} \times X$ to X. This is an example of a **topological vector space**, a set with a topology and a vector space structure such that vector addition and scalar multiplication are continuous maps.

7.26. Using the characterization of compact subsets of Euclidean spaces, prove the Tikhonov theorem for finite products of compact subsets of Euclidean spaces.

7.27. Using the characterization of compact metric spaces in terms of sequences, prove the Tikhonov theorem for finite products of compact metric spaces.

7.28 (the Hilbert cube is metrizable). Consider the set $[0,1]^{\mathbb{Z}^+} = \prod_{n\in\mathbb{Z}^+}[0,1]$ where $[0,1] \subset \mathbb{R}$ is equipped with the Euclidean topology. It is the set of all sequences of real numbers in $[0,1]$.

[1] hội rời

(a) With $x = (x_n)_{n\in\mathbb{Z}^+}$ and $y = (y_n)_{n\in\mathbb{Z}^+}$ in $[0,1]^{\mathbb{Z}^+}$, let

$$d(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n}|x_n - y_n|.$$

Check that this is a metric.

(b) Check that convergence in this metric is coordinate-wise convergence.

(c) Independently, check that open sets in this metric are same as open set in the product topology.

(d) Conclude that this metric generates the product topology of $[0,1]^{\mathbb{Z}^+}$. Thus the Hilbert cube is metrizable.

7.29. In $[0,1]^{\mathbb{Z}^+}$ where $[0,1] \subset \mathbb{R}$ is equipped with the Euclidean topology, with $x = (x_n)_{n\in\mathbb{Z}^+}$ and $y = (y_n)_{n\in\mathbb{Z}^+}$, let

$$d_\infty(x,y) = \sup\{|x_n - y_n| \mid n \in \mathbb{Z}^+\}.$$

Check that this is a metric. Is the topology generated by this metric the product topology?

7.30. * Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. A polynomial in n variables on $\mathbb{F}$ is a function from $\mathbb{F}^n$ to $\mathbb{F}$ that is a finite sum of terms of the form $ax_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$, where a, $x_i \in \mathbb{F}$ and $m_i \in \mathbb{N}$. Let P be the set of all polynomials in n variables on $\mathbb{F}$.

If $S \subset P$ then define $Z(S)$ to be the set of all common zeros of all polynomials in S, thus $Z(S) = \{x \in \mathbb{F}^n \mid \forall p \in S,\ p(x) = 0\}$. Such a set is called an **algebraic set**.

(a) Show that if we define that a subset of $\mathbb{F}^n$ is closed if it is algebraic, then this gives a topology on $\mathbb{F}^n$, called the **Zariski topology**.

(b) Show that the Zariski topology on $\mathbb{F}$ is exactly the finite complement topology.

(c) Show that if both $\mathbb{F}$ and $\mathbb{F}^n$ have the Zariski topology then all polynomials on $\mathbb{F}^n$ are continuous.

(d) Is the Zariski topology on $\mathbb{F}^n$ the product topology?

The Zariski topology is used in Algebraic Geometry.

7.31. The set of $n \times n$-matrix with real coefficients, denoted by $M(n,\mathbb{R})$, could be naturally considered as the Euclidean space $\mathbb{R}^{n^2}$ by considering entries of a matrix as coordinates, via the map

$$(a_{i,j}) \longmapsto (a_{1,1}, a_{2,1}, \ldots, a_{n,1}, a_{1,2}, a_{2,2}, \ldots, a_{n,2}, a_{1,3}, \ldots, a_{n-1,n}, a_{n,n}).$$

Let $\mathrm{GL}(n,\mathbb{R})$ be the set of all invertible $n \times n$-matrices with real coefficients.

(a) Show that taking product of two matrices is a continuous map on $\mathrm{GL}(n,\mathbb{R}) \times \mathrm{GL}(n,\mathbb{R})$.

(b) Show that taking inverse of a matrix is a continuous map on $\mathrm{GL}(n,\mathbb{R})$.

(c) Show that $\mathrm{GL}(n,\mathbb{R})$ is a topological group (see 7.32). It is called the **General Linear Group**.

7.32. A set with both a group structure and a topology such that the group operations are continuous is called a **topological group**.

(a) Show that the circle S^1 identified with the set of unit complex numbers under complex number multiplication and the Euclidean topology is a topological group. Which is the identity element in this group?

(b) Let G be a topological group. Let A and B be two subset of G. Let $AB = \{ab \mid a \in A, b \in B\}$. Show that if A is open then AB is open. Show that if A and B are compact then AB is compact.

(c) Let e be the identity element of G. Show that if $\{e\}$ is a closed set in G then G is a T_1-space.

(d) Let U be an open set containing e. Show that there is an open set V such that $e \in V \subset U$ and $V \cdot V \subset U$. Let $W = V \cap V^{-1}$. Suppose $x \notin U$, show that $xW \cap W = \emptyset$. Deduce that a T_1 topological group is T_2.

(e) Show that if H is a subgroup of G then $\overline{H}$ is also a subgroup of G.

For more see e.g. [Mor24, Appendix 5].

8 Quotient space

We study gluing spaces to form new spaces.

Example. Gluing the two endpoints of a line segment together we get a circle, see Fig. 8.1.

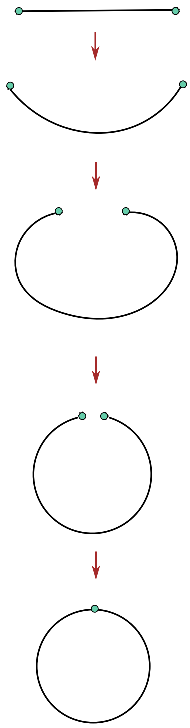


Figure 8.1: Gluing the two end-points of a line segment gives a circle.

Mathematically, gluing elements mean to identify those elements as one element in a new set. Identifying two elements as one means the two elements are related by a relation which should be reflexive, symmetric, and transitive, that is, an equivalence relation. For a set X and an equivalence relation $\sim$ on X, taking the quotient set $X/\sim$ is exactly what we mean by gluing.

Example. In a special case, given a subspace A of X we can think of a space obtained from X by collapsing the whole subspace A into one point. This is the quotient space $X/\sim$ with $\sim$ is the equivalence relation on X where $x \sim x$ if $x \in X$ and $x \sim y$ if $x, y \in A$. The quotient set $X/\sim$ is often written as X/A.

The gluing map is exactly the quotient map bringing each element to its equivalence class,

$$\begin{aligned} p : X &\to X/\sim \\ x &\mapsto [x]. \end{aligned}$$

We want the gluing operation to be continuous. That means when X is a topological space we want to equip the quotient set $X/\sim$ with a topology such that the quotient map p is continuous.

Definition. *The finest topology on $X/\sim$ such that p is continuous is called the* **quotient space** *of X by the equivalence relation $\sim$. Namely, a subset U of $X/\sim$ is open in the quotient topology if and only if the preimage $p^{-1}(U) = \{x \in X \mid [x] \in U\}$ is open in X.*

This is a special case of topologies generated by maps, see 9.1.

8.2 Proposition. *Let $X/\sim$ be a quotient space with quotient map p and let Y be a topological space. A map $f : X/\sim \longrightarrow Y$ is continuous if and only if $f \circ p$ is continuous.*

$$\begin{tikzcd} X \arrow[r, "p"] \arrow[dr, "f\circ p"'] & X/\sim \arrow[d, "f"] \\ & Y \end{tikzcd}$$

Proof. The map $f \circ p$ is continuous if and only if for each open subset U of Y, the set $(f \circ p)^{-1}(U) = p^{-1}(f^{-1}(U))$ is open in X. The latter statement is equivalent to that $f^{-1}(U)$ is open for every U, that is, f is continuous.

This is actually a special case of a general result for topologies generated by maps, see 9.15. □

The following result provides our main tool for identifying quotient spaces:

8.3 Theorem. *Consider $X/\sim$ where $\sim$ is an equivalence relation on X be a quotient space and let $f : X \to Y$ be continuous, onto, and such that $f(x_1) = f(x_2)$ if and only if $x_1 \sim x_2$, then f induces a continuous bijective map from $X/\sim$ onto Y. If X is compact and Y is Hausdorff then f induces a homeomorphism from $X/\sim$ onto Y.*

Proof. Define $h : X/\sim \longrightarrow Y$ by $h([x]) = f(x)$, then h is well-defined, onto, and is injective. Note that $f = h \circ p$.

$$\begin{tikzcd} X \arrow[r, "p"] \arrow[dr, "f"'] & X/\sim \arrow[d, "h"] \\ & Y \end{tikzcd}$$

By 8.2 h is continuous. If X is compact and Y is Hausdorff then by 6.14, h is a homeomorphism. □

Common quotient spaces

8.4 Example (gluing the two end-points of a line segment gives a circle). See Fig. 8.1. Consider the interval $[0, 1]$ on the Euclidean real number line. Take the minimal equivalence relation $\sim$ for which $0 \sim 1$, namely we only need to add $\forall x \in [0,1], x \sim x$. We often denote the quotient space by this relation by $[0,1]/0 \sim 1$. Consider the following diagram:

$$\begin{tikzcd}
{[0,1]} \arrow[r, "p"] \arrow[dr, "f"'] & {[0,1]/0 \sim 1} \arrow[d, "h"] \\
 & S^1
\end{tikzcd}$$

where f is the map $t \mapsto (\cos(2\pi t), \sin(2\pi t))$. The map f is continuous, onto, and it fails to be injective precisely at $t = 0$ and $t = 1$. Since in the quotient set 0 and 1 are identified, the induced map h on the quotient set becomes a bijection. Theorem 8.3 allows us to check that h is a homeomorphism, noting that $[0,1]/0 \sim 1 = p([0,1])$ is compact since it is the continuous image of a compact space. Thus $[0,1]/0 \sim 1$ is homeomorphic to S^1.

8.5 Example (gluing a pair of opposite edges of a square gives a cylinder). See Fig. 8.6. Let $X = [0,1] \times [0,1]/\sim$ where $(0,t) \sim (1,t)$ for all $0 \le t \le 1$. The map $(s,t) \mapsto f(s,t) = (\cos(2\pi s), \sin(2\pi s), t)$ from $[0,1] \times [0,1]$ to the cylinder $S^1 \times [0,1] \subset \mathbb{R}^3$, is continuous, onto, and fails to be injective precisely at the pairs of points $(0,t)$ and $(1,t)$ for $0 \le t \le 1$.

$$\begin{tikzcd}
{[0,1] \times [0,1]} \arrow[r, "p"] \arrow[dr, "f"'] & {[0,1] \times [0,1]/(0,t) \sim (1,t)} \arrow[d, "\tilde{f}"] \\
 & {S^1 \times [0,1]}
\end{tikzcd}$$

By 8.3, the map f induces a homeomorphism $\tilde{f}$ from X to $S^1 \times [0,1]$.

Figure 8.6: This space is homeomorphic to the cylinder.

8.7 Example (gluing opposite edges of a square gives a torus). The torus [1], denoted by T^2, is defined to be the quotient space $[0,1] \times [0,1]/\sim$ where $(s,0) \sim (s,1)$ and $(0,t) \sim (1,t)$ for all $0 \le s,t \le 1$, see Fig. 8.8.

[1]mặt xuyến

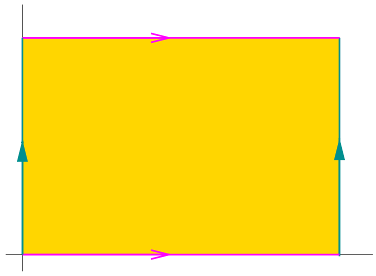


Figure 8.8: The torus.

A subspace of $\mathbb{R}^3$ homeomorphic to the torus can be obtained as the surface of revolution obtained by revolving a circle around a line not intersecting it, see Fig. 8.9.

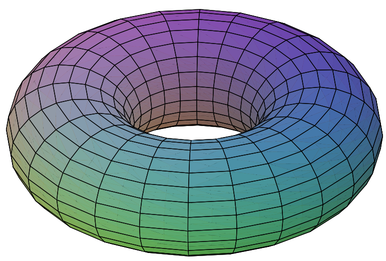


Figure 8.9: The torus embedded in $\mathbb{R}^3$.

In details, suppose that the circle is on the Oyz-plane, the center is on the y-axis and the axis for the rotation is the z-axis, let S be the surface of revolution, see Fig. 8.10.

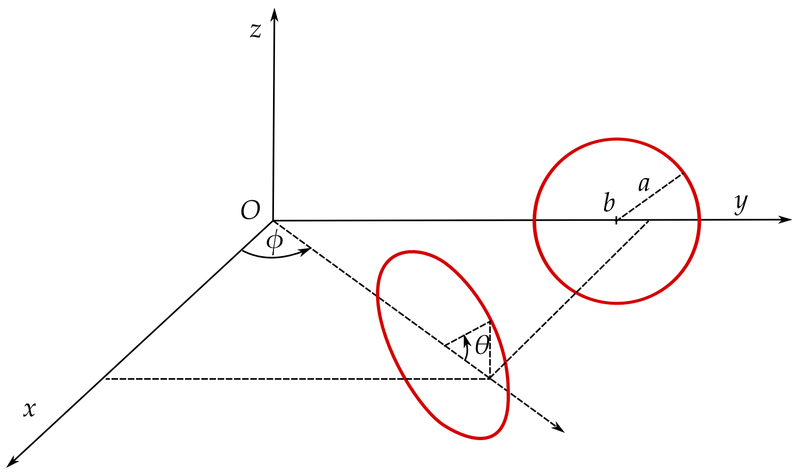


Figure 8.10: Let a be the radius of the circle, b be the distance from the center of the circle to O, this surface has equations $(x, y, z) = ((b + a\cos\theta)\cos\phi, (b + a\cos\theta)\sin\phi, a\sin\theta)$, or $\left(\sqrt{x^2+y^2} - b\right)^2 + z^2 = a^2$.

Consider the maps

$$\begin{tikzcd}
{[0,2\pi]^2} \arrow[r, "p"] \arrow[dr, "f"'] & T^2 \arrow[d, "\tilde{f}"] \\
& S
\end{tikzcd}$$

where

$$f(\phi, \theta) = ((b + a\cos\theta)\cos\phi, (b + a\cos\theta)\sin\phi, a\sin\theta), a < b.$$

By 8.3, f induces a homeomorphism $\tilde{f}$. Thus the torus is homeomorphic to S. We also say that the torus can be embedded in $\mathbb{R}^3$ as the surface S, in the sense of 3.5.

The torus can also be described as the quotient of the Euclidean plane by the relation $\forall m, n \in \mathbb{Z}, (x, y) \sim (x + m, y + n)$, see Problem 8.54 and Fig. 8.11.

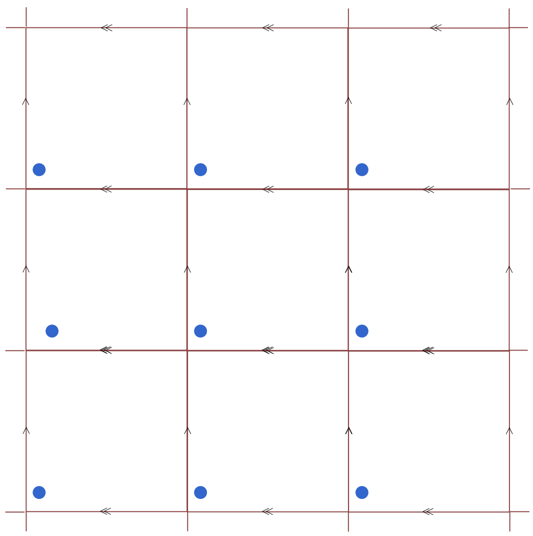


Figure 8.11: The torus as a quotient of the plane.

8.12 Example (Mobius band). Gluing a pair of opposite edges of a square in opposite directions gives the Mobius band. More precisely the **Mobius band** [1] is $X = [0,1] \times [0,1]/\sim$ where $(0, t) \sim (1, 1 - t)$ for all $0 \le t \le 1$. [2]

[1] dải, lá, mặt Mobius

[2] Möbius or Moebius are other spellings.

Figure 8.13: The Mobius band.

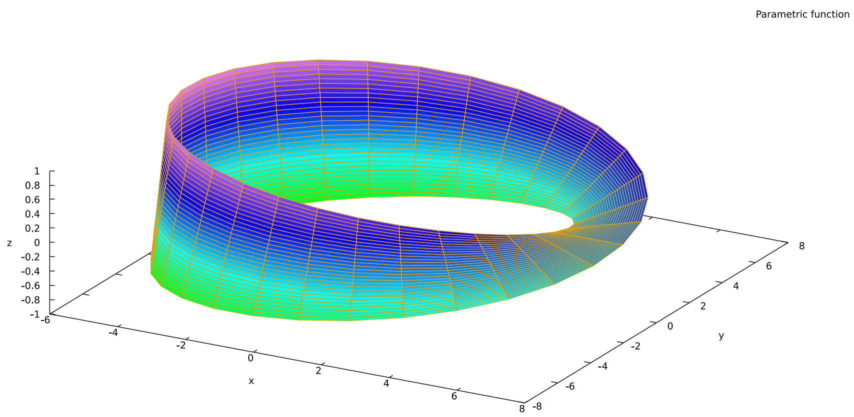


Figure 8.14: The Mobius band embedded in $\mathbb{R}^3$.

The Mobius band could be embedded in $\mathbb{R}^3$. It is homeomorphic to a subspace of $\mathbb{R}^3$ obtained by rotating a straight segment around the z-axis while also turning that segment "up side down". The embedding can be induced by the map (see Fig. 8.15)

$$(s,t) \mapsto ((a + t\cos(s/2))\cos s, (a + t\cos(s/2))\sin s, t\sin(s/2)),$$

with $0 \le s \le 2\pi$ and $-1 \le t \le 1$, where $a > 1$.

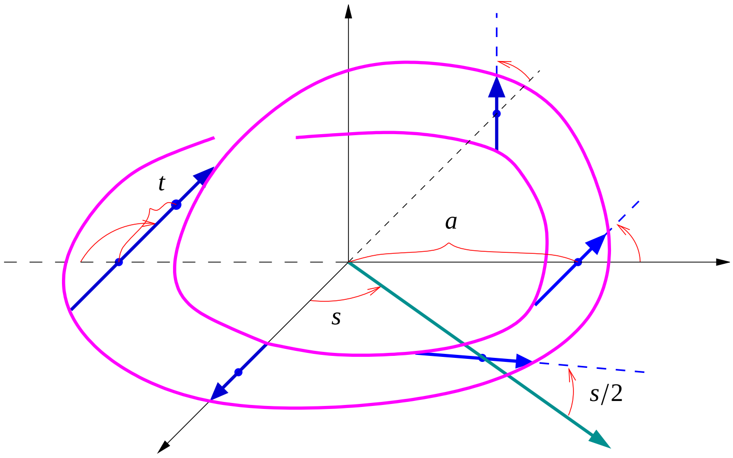


Figure 8.15: An embedding of the Mobius band in $\mathbb{R}^3$.

The Mobius band is famous as an example of **un-orientable** surfaces [1]. It is also **one-sided** [2]. A proof is available at 25.5.

Example (Klein bottle). Identifying one pair of opposite edges of a square and the other pair in opposite directions gives a topological space called the Klein bottle. The **Klein bottle** is the topological space $[0,1]\times[0,1]/\sim$ with $(0,t)\sim(1,t)$ and $(s,0)\sim(1-s,1)$.

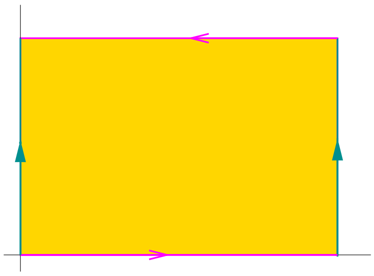


Figure 8.16: The Klein bottle.

It is known that the Klein bottle cannot be embedded in $\mathbb{R}^3$. Figure 8.17 does not present an embedded Klein bottle in $\mathbb{R}^3$, instead only an **immersed** Klein bottle in $\mathbb{R}^3$. An **immersion** is a local embedding, more concisely, $f: X\to Y$ is an immersion if each point in X has a neighborhood U such that $f|_U : U \to f(U)$ is a homeomorphism. Intuitively, an immersion is only required to be bijective locally, not globally, so it allows self-intersections [3].

Example (projective plane). Identifying opposite points on the boundary of a disk gives the projective plane. The **projective plane** [4] $\mathbb{R}\mathrm{P}^2$ is the topological space $D^2/(\forall x\in\partial D^n, x\sim -x)$.

[1] mặt không định hướng được
[2] một phía
[3] tự cắt
[4] mặt phẳng xạ ảnh

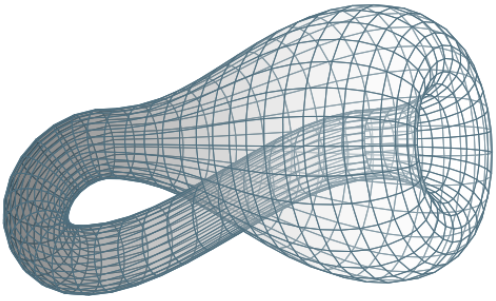


Figure 8.17: The Klein bottle immersed in $\mathbb{R}^3$.

The projective plane cannot be embedded in $\mathbb{R}^3$. This means that there is no subspace of $\mathbb{R}^3$ that is homeomorphic to $\mathbb{R}P^2$. Hence we could not draw it correctly. However the projective plane can be embedded in $\mathbb{R}^4$, see 8.48.

More generally, identifying opposite points (they are called **antipodal points**) on the boundary of D^n gives us **the projective space** [1] $\mathbb{R}P^n$.

Example (three-dimensional torus). Identifying opposite faces of the cube $[0,1]^3$ by $(x,y,0) \sim (x,y,1)$, $(x,0,z) \sim (x,1,z)$, $(0,y,z) \sim (1,y,z)$ we get a topological space called the **three-dimensional torus**. It can also be described, similarly to the two-dimensional torus, as the quotient space of $\mathbb{R}^3$ by the relation $\forall m,n,p \in \mathbb{Z}, (x,y,z) \sim (x+m, y+n, z+p)$. The second description is demonstrated in Fig. 8.18 [2], for more discussions see [Wee02].

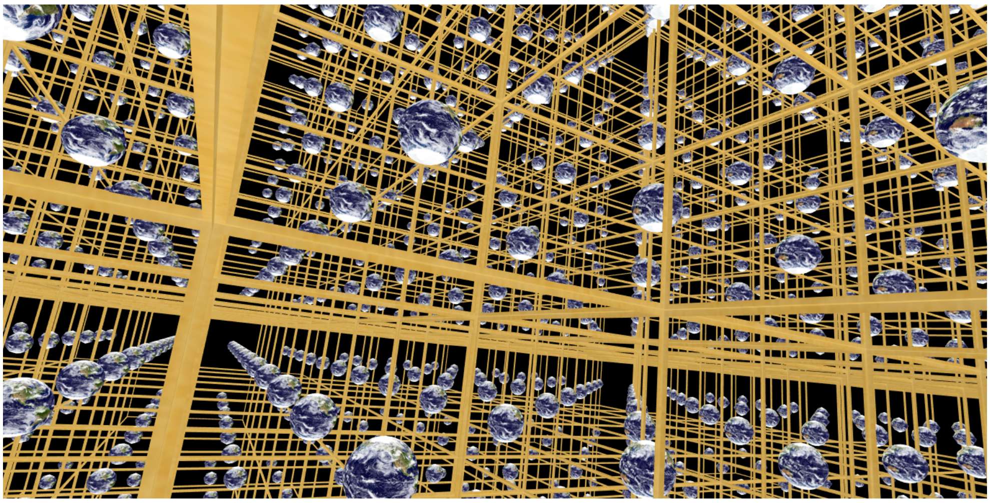


Figure 8.18: An object in the three-dimensional torus can see infinitely many copies of itself.

8.19 Example (wedge sum). Let $(X_i)_{i\in I}$ be a collection of spaces together with a collection of points $x_i \in X_i$. The quotient space of the disjoint union of this collection of spaces (see 7.24) by the relation such that all the points x_i are

[1] không gian xạ ảnh

[2] Figure produced by a computer program available at `http://geometrygames.org`

identified, $\bigsqcup_{i\in I} X_i/(x_i)_{i\in I}$, is called the **wedge sum** [1] of $(X_i)_{i\in I}$ with respect to the points $(x_i)_{i\in I}$, denoted by $\vee_{i\in I}(X_i, x_i)$.

Example (bouquet of circles). A wedge sum of circles $S^1 \vee S^1 \vee \cdots \vee S^1$ (which does not depend on how the identified points are chosen, see Problem 8.20) is called a **bouquet of circles** [2].

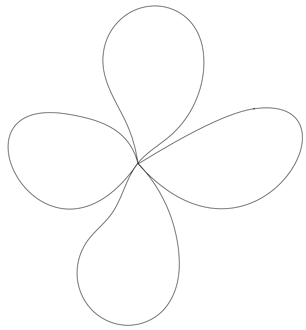


Figure 8.20: $S^1 \vee S^1 \vee S^1 \vee S^1$.

Cut and paste

A quotient space can be formed in different orders of identifications, where parts of the space are consecutively cut and glued, often called "cut and paste". The method is demonstrated in the following examples.

Example (gluing two Mobius bands along their boundaries gives the Klein bottle). See Fig. 8.21. The boundary of a Mobius band is homeomorphic to a circle, this is verified in 8.23. The gluing of two Mobius bands along their boundaries can be done in steps. Together with an appropirate cut, the quotient space is shown to be homeomorphic to the Klein bottle. [3]

Example (gluing a disk to the Mobius band gives the projective plane). See Fig. 8.22, where it is shown that the Mobius band is homeomorphic to the projective plane with a disk deleted.

Problems

8.23. Show that gluing two pairs of endpoints of two line segments gives a circle. Precisely, $([0,1] \cup [2,3])/0 \sim 2, 1 \sim 3$ is homeomorphic to S^1.

[1] tổng nêm, chèn

[2] chùm đường tròn

[3] There is a limerick (humorous poem):
A mathematician named Klein
Thought the Mobius band was divine
Said he, "If you glue
The edges of two,
You'll get a weird bottle like mine."

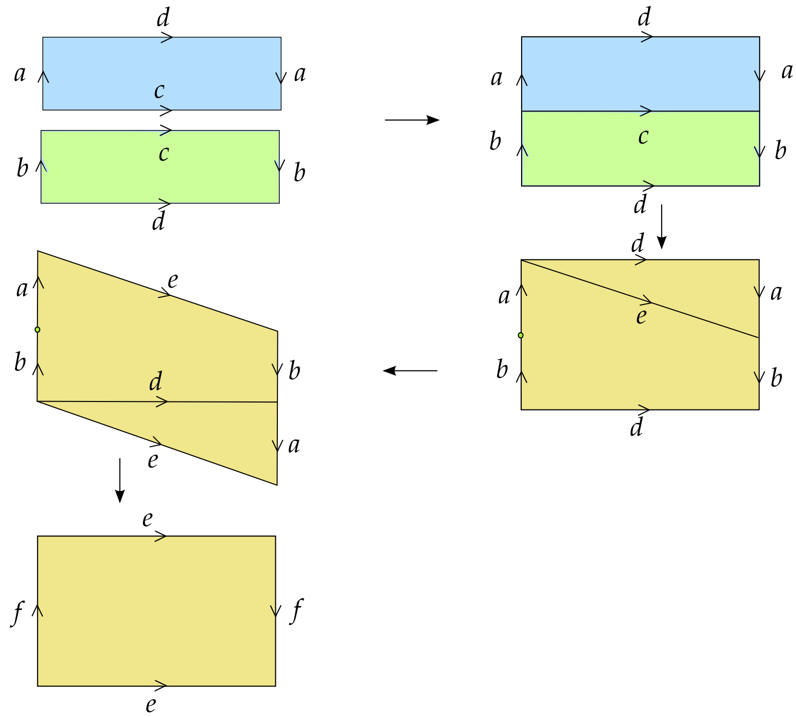


Figure 8.21: Gluing two Mobius bands along their boundaries gives the Klein bottle.

8.24. Show that gluing disks along their boundaries gives a sphere. Precisely, $((D^2 \times \{1\}) \cup (D^2 \times \{2\}))/\forall x \in S^1, (x,1) \sim (x,2)$ is homeomorphic to S^2.

8.25. ✓ Show that the torus T^2 is homeomorphic to $S^1 \times S^1$.

8.26. Let I be the rectangle $[0,1] \times [0,2\pi]$ in the Euclidean plane. Let $\sim$ be the equivalence relation on I satisfying all of the requirements below:

(a) $\forall r \in [0,1], (r,0) \sim (r,2\pi)$,

(b) $\forall \theta \in [0,2\pi], (0,\theta) \sim (0,0)$.

Let $f : I \to D^2$ be the polar coordinates on the plane, given by $f(r,\theta) = (r\cos\theta, r\sin\theta)$. Show that f induces a homeomorphism from $I/\sim$ onto the disk D^2. Can we explain this visually?

8.27. Let I be the rectangle $[0,2\pi] \times [0,\pi]$ in the Euclidean plane. Let $\sim$ be the equivalence relation on I satisfying all of the requirements below:

(a) $\forall \phi \in [0,\pi], (0,\phi) \sim (2\pi,\phi)$,

(b) $\forall \theta \in [0,2\pi], (\theta,0) \sim (0,0)$,

(c) $\forall \theta \in [0,2\pi], (\theta,\pi) \sim (0,\pi)$.

Let $f : I \to S^2$ be the spherical coordinates on the sphere, given by $f(\theta,\phi) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$. Show that f induces a homeomorphism from $I/\sim$ onto S^2. Can it be explained visually?

8.28. Show that gluing the boundary circle of a disk altogether gives a sphere, that is $D^2/\partial D^2$ is homeomorphic to S^2, see Fig. 8.29.

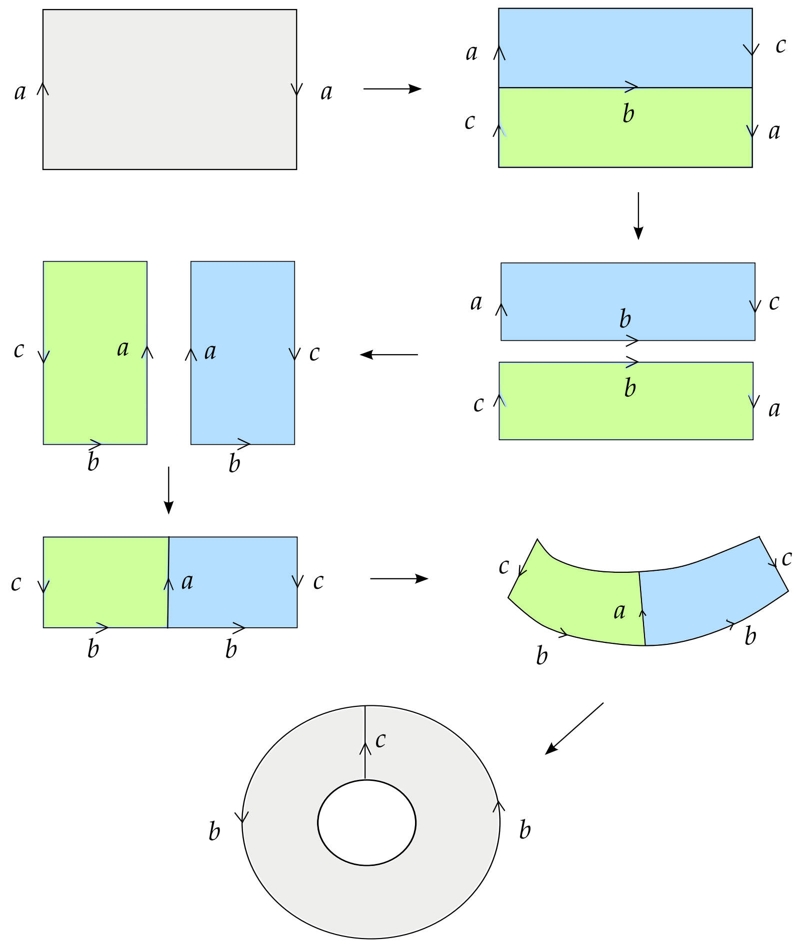


Figure 8.22: The Mobius band is the projective plane minus a disk.

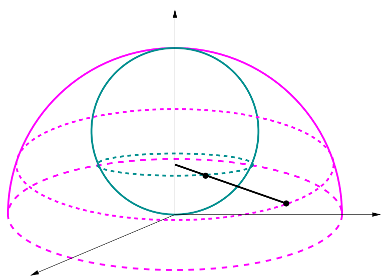


Figure 8.29: A map from D^2 onto S^2 which becomes injective on $D^2/\partial D^2$.

8.30. Show that $S^1 \vee S^1$ does not depend on how the identified points are chosen, and is homeomorphic to the figure-8, namely, this subspace of the Euclidean plane $\{(x, y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 = 1\} \cup \{(x, y) \in \mathbb{R}^2 \mid (x+1)^2 + y^2 = 1\}$, see Fig. 8.31.

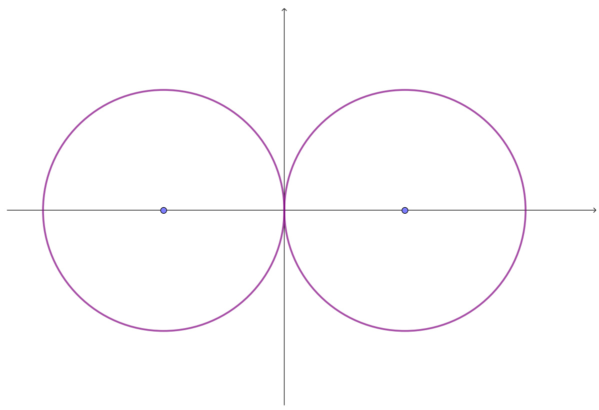


Figure 8.31: $S^1 \vee S^1$ is homeomorphic to the figure-8.

8.32. ✓ By cut-and-paste, show that the space in Fig. 8.33 is homeomorphic to the Mobius band.

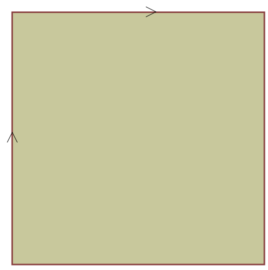


Figure 8.33: Another Mobius band.

8.34. ✓ By cut-and-paste, show that the two spaces in Fig. 8.35 are homeomorphic, one of the two is the Klein bottle.

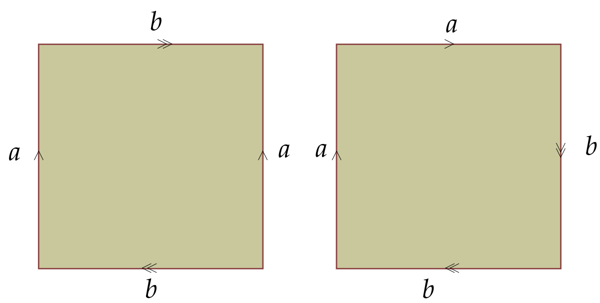


Figure 8.35: Two Klein bottles.

8.36. ✓ What do we obtain after we cut a Mobius band along its middle circle? Try it with an experiment. Mathematically, to cut a subset S from a space X means to delete S from X, the resulting space is the subspace $X \setminus S$. In Fig. 8.37 the curve CC' is deleted.

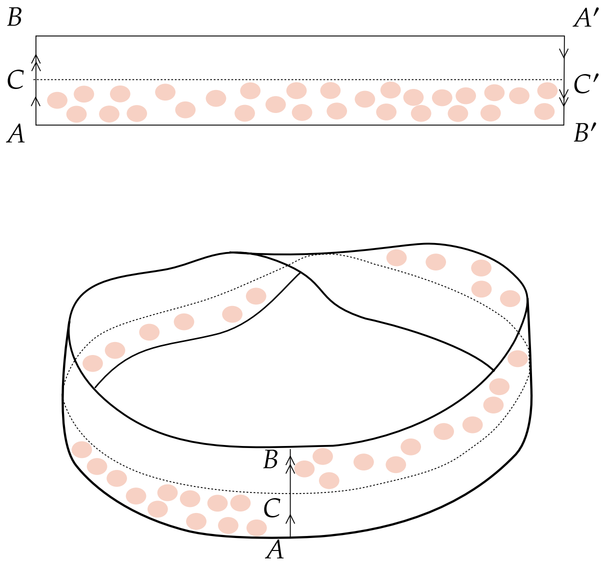


Figure 8.37: Cutting a Mobius band along the middle circle.

8.38. To obtain a physical Mobius band we can take a long rectangular piece of paper, twist one side once (an angle of 180 degree), then glue to the opposite side. What happen if we twist twice? What happen if we twist many times? Do a physical experiment and a computer experiment. See Fig. 8.39.

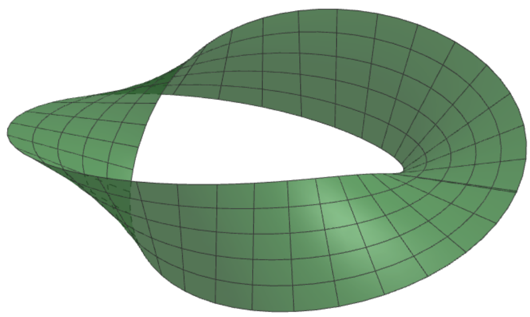


Figure 8.39: A rectangle twisted twice then glued.

8.40. ✓ Following the description of the torus a the quotient space of a rectangle, a line with slope 2/3 in the rectangle after the quotient will be a closed curve on the torus that goes around the torus 2 times in one direction and 3 times in another direction, see Fig. 8.41.

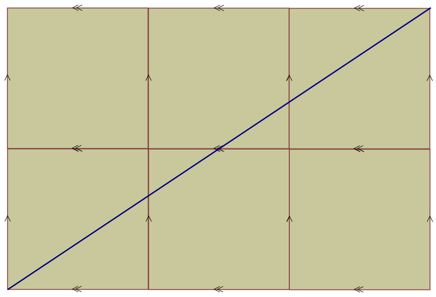


Figure 8.41: The trefoil knot on the torus.

From this it is not difficult to obtain a parametrization of an embedding of that space in $\mathbb{R}^3$ as for example

$$((2+\cos(t/2))\cos(t/3), (2+\cos(t/2))\sin(t/3), \sin(t/2)),\ 0 \le t \le 12\pi,$$

see Fig. 8.42, from which we can see why this space is often called the **trefoil knot**, and compare with Fig. 8.10.

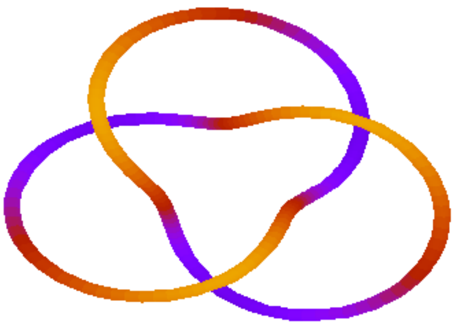


Figure 8.42: The trefoil knot in $\mathbb{R}^3$.

Show that the trefoil knot is homeomorphic to the circle S^1.

8.43. In general, the image of a simple closed curve in a Hausdorff topological space X is called a **knot** [1] in X. Thus a knot is the subspace $\gamma([0,1])$ where $\gamma : [0,1] \to X$ is a continuous map such that $\gamma(0) = \gamma(1)$ and $\gamma|_{[0,1)}$ is injective, see Fig. 8.44. Show that any knot is homeomorphic to the circle.

8.45. Show that the projective space $\mathbb{RP}^1$ is homeomorphic to S^1.

8.46. The one-point compactification of the open Mobius band (the Mobius band without the boundary circle) is the projective space $\mathbb{RP}^2$.

8.47. * Show that identifying antipodal boundary points of D^n is equivalent to identifying antipodal points of S^n. In other words, the projective space $\mathbb{RP}^n$ is homeomorphic to $S^n/x \sim -x$.

[1] nút

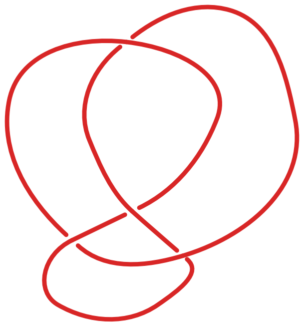


Figure 8.44: A knot is the image of a simple closed curve. This knot is called the**figure-8 knot**, from its shape.

8.48. Consider the projective plane $\mathbb{RP}^2$ as the a quotient space of the sphere $S^2/x \sim -x$, see 8.47. Show that the map

$$\begin{array}{cccc} f: & S^2 & \to & \mathbb{R}^4 \\ & (x,y,z) & \mapsto & (x^2-y^2, xy, yz, zx) \end{array}$$

induces an imbedding of $\mathbb{RP}^2$ into $\mathbb{R}^4$.

8.49. Show that if X has one of the properties connected, path-connected, compact, then so is the quotient space $X/\sim$.

8.50. Show that in order for the quotient space $X/\sim$ to be a Hausdorff space, a necessary condition is that each equivalence class $[x]$ must be a closed subset of X. Is this condition sufficient?

8.51. Suppose $\sim$ is an equivalence relation on X and $f: X \to Y$ is a continuous map satisfying if $x_1 \sim x_2$ then $f(x_1) = f(x_2)$. Then there exists a unique continuous map $g: X/\sim \to Y$ such that $f = g \circ p$.

$$\begin{tikzcd} X \arrow[r, "p"] \arrow[dr, "f=g\circ p"'] & X/\sim \arrow[d, "g"] \\ & Y \end{tikzcd}$$

8.52. Suppose $f: X \to Y$ is continuous and onto and $\sim$ is an equivalence relation on X given by $x_1 \sim x_2$ if and only if $f(x_1) = f(x_2)$. If f is an open map then f induces a homeomorphism from $X/\sim$ onto Y.

8.53. On the Euclidean $\mathbb{R}$ define $x \sim y$ if $x - y \in \mathbb{Z}$. Show that $\mathbb{R}/\sim$ is homeomorphic to S^1. The space $\mathbb{R}/\sim$ is also described as "the quotient of $\mathbb{R}$ by the action of the group $\mathbb{Z}$", written briefly as $\mathbb{R}/\mathbb{Z} = S^1$.

8.54. On the Euclidean $\mathbb{R}^2$, define $(x_1, y_1) \sim (x_2, y_2)$ if $(x_1 - x_2, y_1 - y_2) \in \mathbb{Z}\times\mathbb{Z}$. Show that $\mathbb{R}^2/\sim$ is homeomorphic to the torus, written briefly as $\mathbb{R}^2/\mathbb{Z}^2 = T^2$.

8.55. In combinatorics, a **graph** G is usually defined as a triple (V, E, f), where the set V is called the set of vertices, the set E is called the set of edges, and the map $f: E \to \{\{u, v\} \mid u \in V, v \in V\}$ gives the vertices of each edge ([MT01], [GT01]).

This allows multiple edges between two vertices and loops at one vertex (also called multigraph).

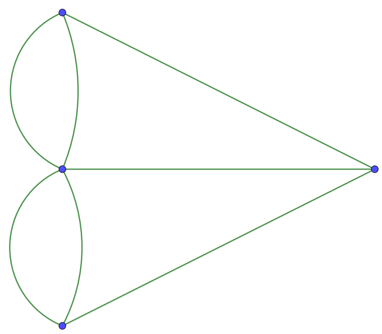


Figure 8.56: A topological graph representing the seven bridges in the problem "Seven bridges of Konigsberg" in Fig 0.1 on page 1.

Below we consider an approach to realize a combinatorial graph as a topological space, as in [Hat01, p. 83]. Given a combinatorial graph $G = (V, E, f)$. Let V have the discrete topology. For each $e \in E$, let I_e be the interval $[0, 1]$ with the Euclidean topology. Consider the disjoint union

$$V \sqcup \left(\bigsqcup_{e \in E} I_e \right).$$

Take the quotient space by gluing $\partial I_e = \{0, 1\}$ to the set of vertices $f(e)$ of e:

$$\tilde{G} = \left[V \sqcup \left(\bigsqcup_{e \in E} I_e \right) \right] / \left[\partial I_e \sim f(e) \right].$$

The topological space $\tilde{G}$ is our topological graph.

Describe a basis for the topology of this topological graph.

8.57. Given a set X and a relation R on X, a non-empty sub set of $X \times X$, show that there is an equivalence relation on X that contains R and is contained in every equivalence relation that contains R, called the **minimal equivalence relation** containing R.

In Example 8.4, we write $[0, 1]/0 \sim 1$ to mean the quotient of the set $[0, 1]$ by the minimal equivalence relation on $[0, 1]$ such that $0 \sim 1$. In this case the minimal equivalence relation is clearly $\{(0, 1), (1, 0), (x, x) \mid x \in [0, 1]\}$.

8.58 (cut and paste). Let X be a topological space and let $A, B \subset X$ such that $A \cup B = X$, A and B are both closed or both open. Let $C = A \cap B$. Let $A \times \{0\}$ and $B \times \{1\}$ have the product topologies. Equip $Y = (A \times \{0\}) \cup (B \times \{1\}) / \forall x \in C, (x, 0) \sim (x, 1)$ with the quotient topology (this is related to 7.24). Roughly, this Y is obtained by cutting X into two parts A and B then glue them back along their common part C. Show that Y is homeomorphic to X.

8.59. * We often visualize the torus as a quotient space in Example 8.7 by two steps: first identify a pair of opposite edges of a rectangle, obtaining a cylinder as in Example 8.5, then identify the opposite circles of the cylinder to get the torus. This is also a process used in cut and paste arguments.

A question can be raised: In quotient spaces, if identifications are carried out in steps rather than simultaneously, will the results be different? More precisely, let R_1 and R_2 be two equivalence relations on a space X and let R be the minimal equivalence relation containing $R_1 \cup R_2$ (see Problem 8.57). On the space X/R_1 define an equivalence relation $\tilde{R}_2$ induced from R_2 by $[x]_{R_1} \sim_{\tilde{R}_2} [y]_{R_1}$ if $x \sim_R y$. Prove that the map

$$\begin{aligned} X/R &\to (X/R_1)/\tilde{R}_2 \\ [x]_R &\mapsto [[x]_{R_1}]_{\tilde{R}_2} \end{aligned}$$

is a homeomorphism. Thus in this sense the results are same.

8.60. Show that if H is a normal subgroup of a topological group G (see 7.32) then the quotient group G/H under the quotient topology is a topological group.

9 Real functions and Spaces of functions

In this section the set $\mathbb{R}$ is assumed to have the Euclidean topology.

Topology generated by maps

In many circumstances we have maps first and we want to build topologies so that these maps become continuous.

In one situation, let (X, τ_X) be a topological space, Y be a set, and $f : X \to Y$ be a map, we want to find a topology on Y such that f is continuous. The requirement for such a topology τ_Y is that if $V \in \tau_Y$ then $f^{-1}(V) \in \tau_X$. The trivial topology on Y is the coarsest topology satisfying that requirement. The collection $\tau_Y = \{V \subset Y \mid f^{-1}(V) \in \tau_X\}$ is the finest topology satisfying that requirement.

In another situation, let X be a set, (Y, τ_Y) be a topological space, and $f : X \to Y$ be a map, we want to find a topology on X such that f is continuous. The requirement for such a topology τ_X is that if $U \in \tau_Y$ then $f^{-1}(U) \in \tau_X$. The discrete topology on X is the finest topology satisfying that requirement. The collection $\tau_X = \{f^{-1}(V) \mid V \in \tau_Y\} = \{U \subset X \mid \exists V \in \tau_Y, U = f^{-1}(V)\}$ is the coarsest topology satisfying that requirement.

In both situations, the topology is called the **topology generated by the map**.

Example. When X is a subset of a topological space Y, then the topology on X as a subspace of Y is exactly the topology generated by the inclusion map $X \hookrightarrow Y$, as noted at 3.2.

Example. For a space X and an equivalence relation $\sim$ on X, the quotient topology on $X/\sim$ is generated by the projection map $p : X \to X/\sim$, $p(x) = [x]$, as seen in Section 8.

Example. Given a real-valued function on a set X, that is a map $f : X \to \mathbb{R}$, where $\mathbb{R}$ has the Euclidean order and metric. Suppose we want the map f to be continuous. The coarsest topology on X for that purpose is generated by the collection of subsets $f^{-1}((a, b))$ for all intervals (a, b) of $\mathbb{R}$.

More generally, we consider the **topology generated by a collection of maps**. .

9.1 Proposition. *Let (X_i, τ_{X_i}) be topological spaces, let Y be a set, and let $f_i : X_i \to Y$, $i \in I$ be a collection of maps. The finest topology on Y such that all maps f_i, $i \in I$ are continuous is $\{V \subset Y \mid \forall i \in I, f_i^{-1}(V) \in \tau_{X_i}\}$. It is the intersection of the topologies generated by each map f_i.*

Let X be a set, let (Y_i, τ_{Y_i}), $i \in I$ be topological spaces, and let $f_i : X \to Y_i$, $i \in I$

be a collection of maps. The coarsest topology on X such that all maps f_i, $i \in I$ are continuous is the topology generated by the collection $\{f_i^{-1}(V) \mid \exists i \in I, V \in \tau_{Y_i}\} = \{U \mid \exists i \in I, \exists V \in \tau_{Y_i}, U = f_i^{-1}(V)\}$. It is the topology generated by the union of the topologies generated by each map f_i.

In both cases, the topology is called the **topology generated by the collection of maps**.

This is Problem 9.13.

Example. The product topology is the topology generated by the projections to component spaces, as seen at 7.4.

9.2 Proposition. *Let (Y_i, τ_{Y_i}) be topological spaces and let X have the topology generated by the collection of maps $\{f_i : X \to Y_i \mid i \in I\}$.*

(a) A net $(x_j)_{j\in J}$ in X converges to a point $x \in X$ if and only if for all $i \in I$, the net $(f_i(x_j))_{j\in J}$ converges to $f_i(x)$.

(b) Let Z be a topological space, and $g : Z \to X$. Then g is continuous if and only if $\forall i \in I$, $f_i \circ g$ is continuous.

This is Problem 9.14.

Example. Propositions 7.5 and 7.6 for the product topology are special cases of 9.2.

Example (the weak topology). On a topological vector space (see 7.25), the topology generated by all continuous linear functionals is called "the weak topology". More precisely, let X be a topological vector space over the field $\mathbb{F}$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, and let X^* be the set of all maps from X to $\mathbb{F}$ which are linear and continuous (X^* is called the dual space of X), then the weak topology on X is the topology generated by X^* – the coarsest topology on X such that all elements of X^* are continuous.

It is immediate that the weak topology of X is weaker (coarser, a subset of) than the original topology of X.

As a special case, on a normed space the weak topology is the coarsest topology such that all linear functionals which are continuous under the norm are still continuous. A sequence $(x_n)_{n\in\mathbb{Z}^+}$ converges weakly to a point x in X if and only if for any $f \in X^*$ the sequence $(f(x_n))_{n\in\mathbb{Z}^+}$ converges to $f(x)$.

This topology plays important roles in Functional Analysis and its applications. Roughly, weaker topologies may allow more convergence and compactness. For more see for examples [Bre11, ch. 3], [Con90, ch. 5].

Topologies on sets of functions

Let X and Y be two topological spaces. We say that a net $(f_i)_{i\in I}$ of maps from X to Y **converges pointwise** to a map $f : X \to Y$ if for each $x \in X$ the net

$(f_i(x))_{i\in I}$ converges to $f(x)$. The topology generated by this convergence (by 5.3) is called the **pointwise convergence topology** [1].

Now we view a map from X to Y as an element of the set $Y^X = \prod_{x\in X} Y$. In this view a map $f: X \to Y$ is an element $f \in Y^X$, and for each $x \in X$ the value $f(x)$ is the x-coordinate of the element f.

Since convergence in the product topology is coordinate-wise convergence (see 7.6), we immediately obtain:

9.3 Proposition (convergence in the product topology is pointwise convergence). *Let $(f_i)_{i\in I}$ be a net of maps from X to Y, which is also a net of points in Y^X. Then the net of maps $(f_i)_{i\in I}$ converges to a map $f: X \to Y$ pointwise if and only if the net of points $(f_i)_{i\in I}$ converges to the point f in the product topology of Y^X.*

Define the **point-open topology** on the set Y^X of functions from X to Y as the topology generated by sets of the form

$$S(x, U) = \{f \in Y^X \mid f(x) \in U\}$$

with $x \in X$ and $U \subset Y$ is open.

9.4 Proposition. *The point-open topology is exactly the pointwise convergence topology.*

Proof. If we denote by p_x the projection map from $Y^X = \prod_{x\in X} Y$ to the x-component Y, then $p_x(f) = f(x)$ for each $f \in Y^X$. With this view $S(x, U) = \{f \in Y^X \mid f(x) \in U\} = p_x^{-1}(U)$. The topology generated by sets of the form $p_x^{-1}(U)$ is precisely the product topology on Y^X, see 7.4. □

Now let Y be a metric space. Recall that a function $f: X \to Y$ is said to be bounded if the set of values $f(X)$ is a bounded subset of Y. We consider the set $B(X, Y)$ of all bounded functions from X to Y. If $f, g \in B(X, Y)$ then we define a metric on $B(X, Y)$ by $d(f, g) = \sup\{d(f(x), g(x)) \mid x \in X\}$, as in Functional Analysis, see [TTV]. The topology generated by this metric is called the **topology of uniform convergence** [2]. If a net $(f_i)_{i\in I}$ converges to f in the metric space $B(X, Y)$ then we say that $(f_i)_{i\in I}$ converges to f **uniformly**.

Similar to results stated for metric spaces in Functional Analysis, we have:

Proposition. *Suppose that $(f_i)_{i\in I}$ converges to f uniformly. Then:*

(a) $(f_i)_{i\in I}$ converges to f pointwise.

(b) If each f_i is continuous then f is continuous.

Proof. The proof is the same as the proof for metric spaces such as in [TTV]. We write down a proof for the second part. Suppose that each f_i is continuous.

[1] tôpô hội tụ điểm
[2] tôpô hội tụ đều

Let $x \in X$, we prove that f is continuous at x. The key step is the following inequalities:

$$\begin{aligned} d(f(x), f(y)) &\leq d(f(x), f_i(x)) + d(f_i(x), f_i(y)) + d(f_i(y), f(y)). \\ &\leq d(f, f_i) + d(f_i(x), f_i(y)) + d(f_i, f). \end{aligned}$$

Given $\epsilon > 0$, fix an $i \in I$ such that $d(f_i, f) < \epsilon$. For this i, there is a neighborhood U of x such that if $y \in U$ then $d(f_i(x), f_i(y)) < \epsilon$. The above inequality implies that for $y \in U$ we have $d(f(x), f(y)) < 3\epsilon$. □

Definition. *Let X and Y be two topological spaces. Let $C(X, Y)$ be the set of all continuous functions from X to Y. The topology generated by all sets of the form*

$$S(A, U) = \{f \in C(X, Y) \mid f(A) \subset U\}$$

where $A \subset X$ is compact and $U \subset Y$ is open is called the **compact-open topology** *on $C(X, Y)$.*

Example. The compact-open topology contains the point-open topology, which is the pointwise convergence topology, see 9.4. Thus the compact-open topology contains the pointwise convergence topology. Convergence in the compact-open topology implies pointwise convergence.

Proposition. *If X is compact and Y is a metric space then on $C(X, Y)$ the compact-open topology is the same as the uniform convergence topology.*

This shows that the compact-open topology is a generalization of the uniform convergence topology to topological spaces.

Proof. In the first direction, given $f \in C(X, Y)$ and $\epsilon > 0$, we show that the ball $B(f, \epsilon) \subset C(X, Y)$ in the uniform metric contains an open neighborhood of f in the compact-open topology.

For each $x \in X$ there is an open set V_x containing x such that $f(V_x) \subset B(f(x), \epsilon/3)$, see Fig. 9.5.

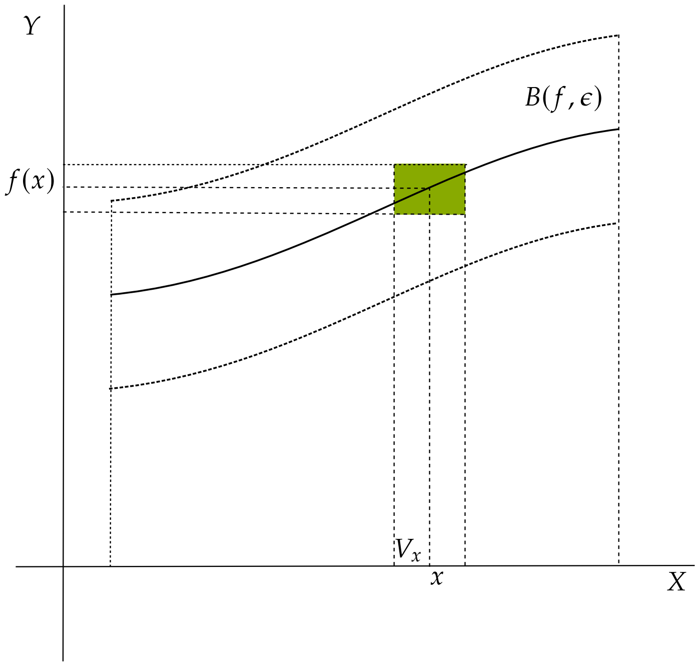


Figure 9.5: $\bigcap_{i=1}^n S(\overline{V}_{x_i}, B(f(x_i), \epsilon/2)) \subset B(f,\epsilon)$.

Since X is compact, there are finitely many x_i, $1 \le i \le n$, such that $\bigcup_{i=1}^n V_{x_i} \supset X$ and $f(V_{x_i}) \subset B(f(x_i), \epsilon/3)$. Using 5.13 we notice that

$$f(\overline{V}_{x_i}) \subset \overline{f(V_{x_i})} \subset \overline{B(f(x_i), \epsilon/3)} \subset B(f(x_i), \epsilon/2),$$

from which we deduce that $f \in \bigcap_{i=1}^n S(\overline{V}_{x_i}, B(f(x_i), \epsilon/2))$. Now if $g \in S(\overline{V}_{x_i}, B(f(x_i), \epsilon/2))$ then for $x \in V_{x_i}$ we have

$$d(g(x), f(x)) \le d(g(x), f(x_i)) + d(f(x_i), f(x)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

implying $g \in B(f,\epsilon)$. Thus $\bigcap_{i=1}^n S(\overline{V}_{x_i}, B(f(x_i), \epsilon/2)) \subset B(f,\epsilon)$.

In the opposite direction, we need to show that every open neighborhood of f in the compact-open topology contains a ball $B(f,\epsilon)$ in the uniform metric. It is sufficient to show that for open neighborhoods of f of the form $S(A,U)$. For each $x \in A$ there is a ball $B(f(x), \epsilon_x) \subset U$. Since $f(A)$ is compact, there are finitely many $x_i \in A$ and $\epsilon_i > 0$, $1 \le i \le n$, such that $B(f(x_i), \epsilon_i) \subset U$ and $\bigcup_{i=1}^n B(f(x_i), \epsilon_i/2) \supset f(A)$. Let $\epsilon = \min\{\epsilon_i/2 \mid 1 \le i \le n\}$, we check that $B(f,\epsilon) \subset S(A,U)$, see Fig. 9.6.

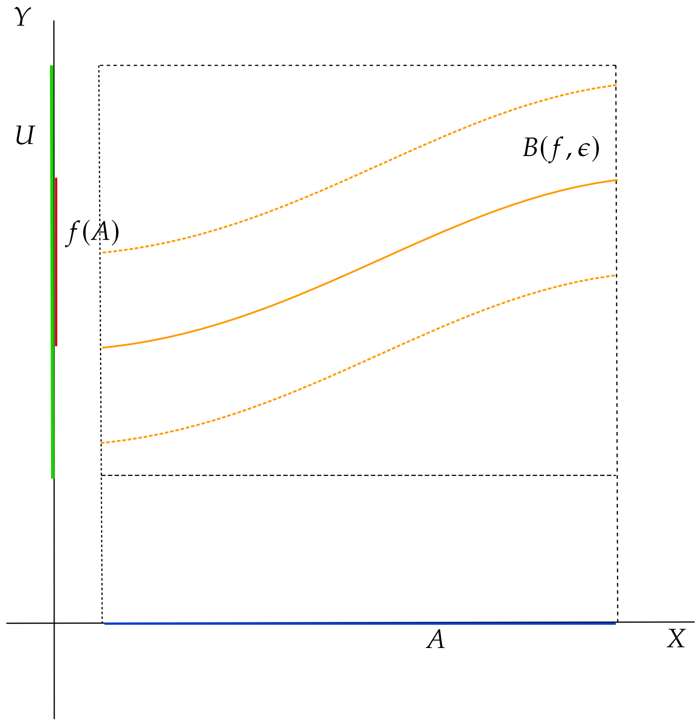


Figure 9.6: $B(f, \epsilon) \subset S(A, U)$.

Suppose that $g \in B(f, \epsilon)$. For each $x \in A$, there is an i such that $f(x) \in B(f(x_i), \epsilon_i/2)$. Then

$$d(g(x), f(x_i)) \leq d(g(x), f(x_i)) + d(f(x_i), f(x)) < \epsilon + \frac{\epsilon_i}{2} \leq \epsilon_i,$$

so $g(x) \in U$. Thus $g \in S(A, U)$, and so $B(f, \epsilon) \subset S(A, U)$. □

Urysohn lemma

9.7 Theorem (Urysohn lemma). *If X is normal, F is closed, U is open, and $F \subset U \subset X$, then there exists a continuous map $f : X \longrightarrow [0, 1]$ such that $f(x) = 0$ on F and $f(x) = 1$ on $X \setminus U$.*

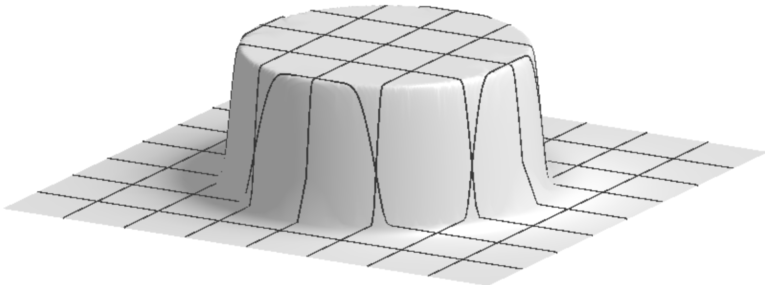
Equivalently, if X is normal, A and B are two disjoint closed subsets of X, then there is a continuous function f from X to $[0, 1]$ such that $f(x) = 0$ on A and $f(x) = 1$ on B.

Thus in a normal space two disjoint closed subsets can be separated by a continuous real function.

Example. For metric spaces we can take

$$f(x) = \frac{d(x, A)}{d(x, A) + d(x, B)}.$$

If we take $g = 1 - f$ we get a continuous function g from X to $[0, 1]$ such that $g|_F = 1$ and $g|_{X \setminus U} = 0$. The graph of g can be visualized as a continuous connection between a highland F at height 1 with the lowland outside U at height 0.



Proof. The proof goes through the following steps:

(a) Construct a family of open sets in the following manner (recalling 5.8):

Let $U_1 = U$.

$n = 0$: $F \subset U_0 \subset \overline{U_0} \subset U_1$.

$n = 1$: $\overline{U_0} \subset U_{\frac{1}{2}} \subset \overline{U_{\frac{1}{2}}} \subset U_1$.

$n = 2$: $\overline{U_0} \subset U_{\frac{1}{4}} \subset \overline{U_{\frac{1}{4}}} \subset U_{\frac{2}{4}} = U_{\frac{1}{2}} \subset \overline{U_{\frac{2}{4}}} \subset U_{\frac{3}{4}} \subset \overline{U_{\frac{3}{4}}} \subset U_{\frac{4}{4}} = U_1$.

$n \in \mathbb{N}$: Inductively,

$$F \subset U_0 \subset \overline{U_0} \subset U_{\frac{1}{2^n}} \subset \overline{U_{\frac{1}{2^n}}} \subset U_{\frac{2}{2^n}} \subset \overline{U_{\frac{2}{2^n}}} \subset U_{\frac{3}{2^n}} \subset \overline{U_{\frac{3}{2^n}}} \subset \cdots \subset$$
$$\subset U_{\frac{2^n-1}{2^n}} \subset \overline{U_{\frac{2^n-1}{2^n}}} \subset U_{\frac{2^n}{2^n}} = U_1.$$

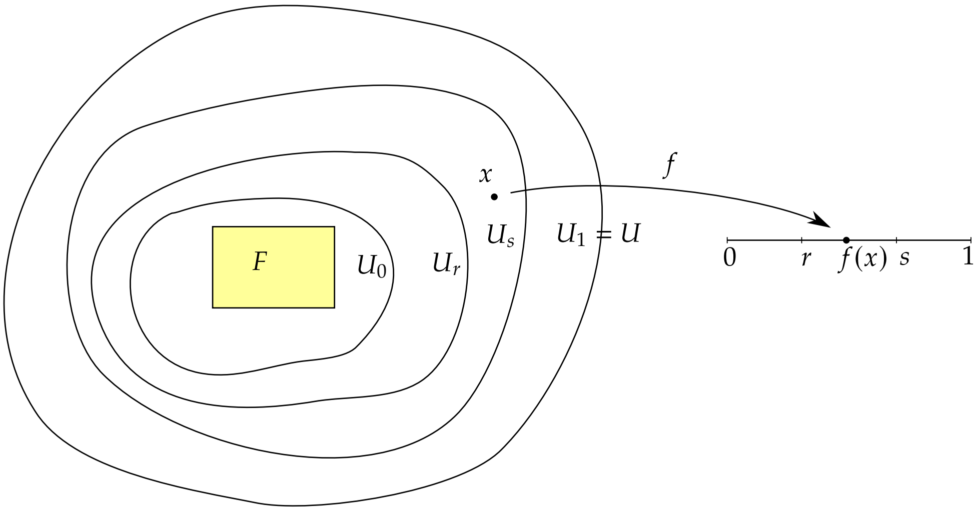
(b) Let $I = \{\frac{m}{2^n} \mid m \in \mathbb{N}, n \in \mathbb{N}, 0 \le m \le 2^n\}$. We have a family of open sets $\{U_r \mid r \in I\}$ having the property $r < s \Rightarrow \overline{U_r} \subset U_s$. We can check that I is dense in $[0,1]$ (this is really the same thing as that any real number in $[0,1]$ can be written in binary form, compare 1.23).

(c) Define $f : X \to [0,1]$,

$$f(x) = \begin{cases} \inf\{r \in I \mid x \in U_r\} & \text{if } x \in U, \\ 1 & \text{if } x \notin U. \end{cases}$$

In this way if $x \in U_r$ then $f(x) \le r$, while if $x \notin U_r$ then $f(x) \ge r$. So $f(x)$ gives the "level" of x on the scale from 0 to 1, while U_r is like a sublevel set of f.

We prove that f is continuous, so f is the function we are looking for. It is enough to prove that sets of the form $\{x \mid f(x) < a\}$ and $\{x \mid f(x) > a\}$ are open.



(d) If $a \leq 1$ then $f(x) < a$ if and only if there is $r \in I$ such that $r < a$ and $x \in U_r$. Thus $\{x \mid f(x) < a\} = \{x \in U_r \mid r \in I, r < a\} = \bigcup_{r \in I, r<a} U_r$ is open.

(e) If $a < 1$ then $f(x) > a$ if and only if there is $r \in I$ such that $r > a$ and $x \notin U_r$. Indeed, if $f(x) > a$ then there is $s > a$ such that $f(x) \geq s$, which in turn implies that if $x \in U_r$ then $r \geq s$, so for $a < r < s$ we have $x \notin U_r$. On the other direction, if $r > a$ and $x \notin U_r$ then $x \in U_s$ implies $s \geq r$, so $f(x) \geq r > a$.
Thus $\{x \mid f(x) > a\} = \{x \in X \setminus U_r \mid r > a\} = \bigcup_{r>a} X \setminus U_r$. Now we show that $\bigcup_{r>a} X \setminus U_r = \bigcup_{r>a} X \setminus \overline{U_r}$, which implies that $\bigcup_{r>a} X \setminus U_r$ is open. Indeed, if $r \in I$ and $r > a$ then there is $s \in I$ such that $r > s > a$. Then $\overline{U_s} \subset U_r$, therefore $X \setminus U_r \subset X \setminus \overline{U_s}$. This implies $\bigcup_{r>a} X \setminus U_r \subset \bigcup_{r>a} X \setminus \overline{U_r}$. The opposite inclusion is true since $X \setminus U_r \supset X \setminus \overline{U_r}$.

□

Partition of unity

An important application of the Urysohn lemma is the existence of **partition of unity** [1]. A partition of unity is a collection of real functions which are non-negative, each function is zero outside a given open set, and the sum of these functions is 1 everywhere. Roughly, a partition of unity is a system of relative weights of each open set at each point (so that total weight at each point is 100%, that is, 1). Figure 9.8 illustrate a simple case.

[1] phân hoạch đơn vị

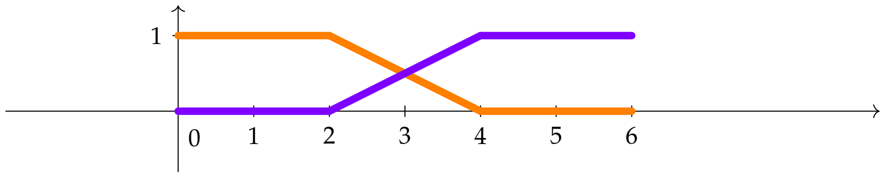


Figure 9.8: A partition of unity for the interval $[0,6]$ corresponding to the two open sets $[0,5)$ and $(1,6]$.

For $f : X \to \mathbb{R}$, the **support** [1] of f, denoted by $\mathrm{supp}(f)$, is defined to be the closure of the subset $\{x \in X \mid f(x) \neq 0\}$.

9.9 Theorem (partition of unity). *Let X be a normal space. Suppose that X has a finite open cover O. Then there is a collection of continuous maps $(f_U : X \to [0,1])_{U \in O}$ such that $\mathrm{supp}(f_U) \subset U$ and for every $x \in X$ we have $\sum_{U \in O} f_U(x) = 1$.*

Proof. From 5.26, there is an open cover $(U''_U)_{U \in O}$ of X such that for each $U \in O$ there is an open U'_U satisfying $U''_U \subset \overline{U''_U} \subset U'_U \subset \overline{U'_u} \subset U$. By Urysohn lemma there is a continuous map $\varphi_U : X \to [0,1]$ such that $\varphi_U|_{U''_U} = 1$ and $\varphi_U|_{X \setminus U'_U} = 0$. This implies $\mathrm{supp}(\varphi_U) \subset \overline{U'_u} \subset U$. For each $x \in X$ there is $U \in O$ such that U''_U contains x, therefore $\varphi_U(x) = 1$. Let

$$f_U = \frac{\varphi_U}{\sum_{U \in O} \varphi_U}.$$

$\square$

Partition of unity allows us to extend some local properties to global ones, by "patching" neighborhoods.

Stone–Cech compactification

Let X be a topological space. Denote by $C(X)$ the set of all bounded continuous functions from X to $\mathbb{R}$ where $\mathbb{R}$ has the Euclidean topology.

By Tikhonov theorem the space $\prod_{f \in C(X)} [\inf f, \sup f]$ is compact. Below we embed the space X into this compact space, hence compactify X.

Define

$$\begin{aligned} \Phi : X &\to \prod_{f \in C(X)} [\inf f, \sup f] \\ x &\mapsto (f(x))_{f \in C(X)}. \end{aligned}$$

Thus for each $x \in X$ and each $f \in C(X)$, the f-coordinate of the point $\Phi(x)$ is $\Phi(x)_f = f(x)$. This means the f-component of Φ is f, i.e. $p_f \circ \Phi = f$, where p_f is the projection to the f-coordinate. This map, where a function is viewed

[1] giá, giá đỡ

as an element of a product space, is used when we discuss the pointwise convergence topology in 9.3.

Notice that the closure $\overline{\Phi(X)}$ is compact. Under additional assumptions we can check that Φ is an embedding.

A space is said to be **completely regular**, also called a $T_{3\frac{1}{2}}$-space, if it is a T_1-space and for each point x and each closed set A with $x \notin A$ there is a map $f \in C(X)$ such that $f(x) = a$ and $f(A) = \{b\}$ where $a \neq b$. Thus in a completely regular space a point and a closed set disjoint from it can be separated by a continuous real function.

9.10 Theorem. *If X is completely regular then $\Phi : X \longrightarrow \Phi(X)$ is a homeomorphism, i.e. Φ is an embedding. In this case $\overline{\Phi(X)}$ is called the* **Stone–Cech compactification** *of X. It is a Hausdorff space.*

Proof. We go through several steps.

(a) Φ is injective: If $x \neq y$ then since X is completely regular there is $f \in C(X)$ such that $f(x) \neq f(y)$, therefore $\Phi(x) \neq \Phi(y)$.

(b) Φ is continuous: Since the f-component of Φ is f, which is continuous, the result follows from 7.5.

(c) Φ^{-1} is continuous: We prove that Φ brings an open set onto an open set. Let U be an open subset of X and let $x \in U$. There is a function $f \in C(X)$ that separates x and $X \setminus U$. In particular there is an interval (a, b) containing $f(x)$ such that $f^{-1}((a, b)) \cap (X \setminus U) = \emptyset$. We have $f^{-1}((a, b)) = (p_f \circ \Phi)^{-1}((a, b)) = \Phi^{-1}(p_f^{-1}((a, b))) \subset U$. Apply Φ to both sides, we get $p_f^{-1}((a, b)) \cap \Phi(X) \subset \Phi(U)$. Since $p_f^{-1}((a, b)) \cap \Phi(X)$ is an open set in $\Phi(X)$ containing $\Phi(x)$, we see that $\Phi(x)$ is an interior point of $\Phi(U)$. We conclude that $\Phi(U)$ is open.

That $\overline{\Phi(X)}$ is a Hausdorff space follows from that Y is a Hausdorff space, by 7.19, and 5.23. □

Theorem. *A bounded continuous real function on a completely regular space has a unique extension to the Stone-Cech compactification of the space.*

More concisely, if X is a completely regular space and $f \in C(X)$ then there is a unique function $\tilde{f} \in C(\overline{\Phi(X)})$ such that $f = \tilde{f} \circ \Phi$.

$$\begin{tikzcd} X \arrow[r, "\Phi"] \arrow[d, "f"'] & \overline{\Phi(X)} \arrow[dl, "\tilde{f}"] \\ \mathbb{R} & \end{tikzcd}$$

Proof. A continuous extension of f, if exists, is unique, by 5.15.

Since $p_f \circ \Phi = f$ the obvious choice for $\tilde{f}$ is the projection p_f restricted to $\overline{\Phi(X)}$. □

* Tiestze extension theorem

We consider another application of Urysohn lemma:

Theorem (Tiestze extension theorem). *Let X be a normal space. Let F be closed in X. Let $f : F \to \mathbb{R}$ be continuous. Then there is a continuous map $g : X \to \mathbb{R}$ such that $g|_F = f$.*

Thus *in a normal space a continuous real function on a closed subspace can be extended continuously to the whole space*.

Proof. First consider the case where f is bounded.

(a) The general case can be reduced to the case when $\inf_F f = 0$ and $\sup_F f = 1$. We will restrict our attention to this case.

(b) By Urysohn lemma, there is a continuous function $g_1 : X \to [0, \frac{1}{3}]$ such that

$$g_1(x) = \begin{cases} 0 & \text{if } x \in f^{-1}([0, \frac{1}{3}]) \\ \frac{1}{3} & \text{if } x \in f^{-1}([\frac{2}{3}, 1]). \end{cases}$$

Let $f_1 = f - g_1$. Then $\sup_X g_1 = \frac{1}{3}$, $\sup_F f_1 = \frac{2}{3}$, and $\inf_F f_1 = 0$.

(c) Inductively, once we have a function $f_n : F \to \mathbb{R}$, for a certain $n \geq 1$ we will obtain a function $g_{n+1} : X \to [0, \frac{1}{3}\left(\frac{2}{3}\right)^n]$ such that

$$g_{n+1}(x) = \begin{cases} 0 & \text{if } x \in f_n^{-1}([0, \frac{1}{3}\left(\frac{2}{3}\right)^n]) \\ \frac{1}{3}\left(\frac{2}{3}\right)^n & \text{if } x \in f_n^{-1}([\left(\frac{2}{3}\right)^{n+1}, \left(\frac{2}{3}\right)^n]). \end{cases}$$

Let $f_{n+1} = f_n - g_{n+1}$. Then $\sup_X g_{n+1} = \frac{1}{3}\left(\frac{2}{3}\right)^n$, $\sup_F f_{n+1} = \left(\frac{2}{3}\right)^{n+1}$, and $\inf_F f_{n+1} = 0$.

(d) The series $\sum_{n=1}^{\infty} g_n$ converges uniformly to a continuous function g.

(e) Since $f_n = f - \sum_{i=1}^{n} g_i$, the series $\sum_{n=1}^{n} g_n|_F$ converges uniformly to f. Therefore $g|_F = f$.

(f) Note that with this construction $\inf_X g = 0$ and $\sup_X g = 1$.

Now consider the case when f is not bounded.

(a) Suppose that f is neither bounded from below nor bounded from above. Let h be a homeomorphism from $(-\infty, \infty)$ to $(0, 1)$. Then the range of $f_1 = h \circ f$ is a subset of $(0, 1)$, therefore it can be extended as in the previous case to a continuous function g_1 such that $\inf_{x \in X} g_1(x) = \inf_{x \in F} f_1(x) = 0$ and $\sup_{x \in X} g_1(x) = \sup_{x \in F} f_1(x) = 1$.

If the range of g_1 includes neither 0 nor 1 then $g = h^{-1} \circ g_1$ will be the desired function.

It may happens that the range of g_1 includes either 0 or 1. In this case let $C = g_1^{-1}(\{0, 1\})$. Note that $C \cap F = \emptyset$. By Urysohn lemma, there is a continuous function $k : X \to [0, 1]$ such that $k|_C = 0$ and $k|_F = 1$. Let $g_2 = kg_1 + (1 - k)\frac{1}{2}$. Then $g_2|_F = g_1|_F$ and the range of g_2 is a subset of $(0, 1)$ ($g_2(x)$ is a certain convex combination of $g_1(x)$ and $\frac{1}{2}$). Then $g = h^{-1} \circ g_2$ will be the desired function.

(b) If f is bounded from below then similarly to the previous case we can use a homeomorphism $h : [a, \infty) \to [0, 1)$, and we let $C = g_1^{-1}(\{1\})$.

The case when f is bounded from above is similar.

□

Problems

9.11. Show that a normal space is completely regular. So: normal ⇒ completely regular ⇒ regular. In other words: $T_4 \Rightarrow T_{3\frac{1}{2}} \Rightarrow T_3$.

9.12. Prove the following version of Urysohn lemma, as stated in [Rud86]. Suppose that X is a locally compact Hausdorff space, V is open in X, $K \subset V$, and K is compact. Then there is a continuous function $f : X \to [0, 1]$ such that $f(x) = 1$ for $x \in K$ and $\operatorname{supp}(f) \subset V$.

9.13. Prove 9.1.

9.14. Prove 9.2.

9.15. Let (X_i, τ_{X_i}) be topological spaces, let Y have the topology generated by the collection of maps $\{f_i : X_i \to Y \mid i \in I\}$. Let Z be a topological space, and $g : Y \to Z$. Then g is continuous if and only if $\forall i \in I$, $g \circ f_i$ is continuous.

9.16 (continuity of functions of two variables). * Let Y, Z be topological spaces and let X be a locally compact Hausdorff space. Recall that it is not true that a map on a product space is continuous if it is continuous on each variable. Prove that a map

$$\begin{aligned} f : X \times Z &\to Y \\ (x, t) &\mapsto f(x, t) \end{aligned}$$

is continuous if and only if the map

$$\begin{aligned} f_t : X &\to Y \\ x &\mapsto f_t(x) = f(x, t) \end{aligned}$$

is continuous for each $t \in Z$ and the map

$$\begin{aligned} Z &\to C(X, Y) \\ t &\mapsto f_t \end{aligned}$$

is continuous with the compact-open topology on $C(X, Y)$.

9.17. Show that the Tiestze extension theorem implies the Urysohn lemma.

9.18. The Tiestze extension theorem is not true without the condition that the set F is closed.

9.19. Show that the Tiestze extension theorem can be extended to maps to the space $\prod_{i \in I} \mathbb{R}$ where $\mathbb{R}$ has the Euclidean topology.

9.20. Let X be a normal space and F be a closed subset of X. Then any continuous map $f : F \to S^n$ can be extended to an open set containing F.

9.21. Any completely regular space is a regular space.

9.22. Prove 9.10 using nets.

9.23. A space is completely regular if and only if it is homeomorphic to a subspace of a compact Hausdorff space. As a corollary, a locally compact Hausdorff space is completely regular.

By 9.23 if a space has a Hausdorff Alexandroff compactification then it also has a Hausdorff Stone–Cech compactification. In a certain sense, for a non-compact space the Alexandroff compactification is the "smallest" Hausdorff compactification of the space and the Stone–Cech compactification is the "largest" one [Mun00, p. 237].

Other topics

Topics for course projects and further study could be chosen from previous sections, in particular from the * problems and from topics in these notes which have not been discussed in class. Below there are several other topics.

Compactness based on nets

We outline here a characterization of compactness in terms of nets and a proof of Tikhonov theorem using nets.

Let I and I' be directed sets, and let $h : I' \to I$ be a map such that

$$\forall k \in I, \exists k' \in I', (i' \geq k' \Rightarrow h(i') \geq k).$$

If $n : I \to X$ is a net then $n \circ h$ is called a subnet of n. The notion of subnet is an extension of the notion of subsequence.

A net $(x_i)_{i \in I}$ is called eventually in $A \subset X$ if there is $j \in I$ such that $i \geq j \Rightarrow x_i \in A$. A net n in X is universal if for any subset A of X either n is eventually in A or n is eventually in $X \setminus A$.

Proposition. *If $f : X \to Y$ is continuous and n is a universal net in X then $f(n)$ is a universal net.*

Proposition. *The following statements are equivalent:*

(a) X is compact.

(b) Every universal net in X is convergent.

(c) Every net in X has a convergent subnet.

Proofs of the above two propositions can be found in [Bre93, p. 23].

Proof of Tikhonov theorem. Let $X = \prod_{i \in I} X_i$ where each X_i is compact. Suppose that $(x_j)_{j \in J}$ is a universal net in X. By 7.6 the net (x_j) is convergent if and only if the projection $(p_i(x_j))$ is convergent for all i. But that is true since $(p_i(x_j))$ is a universal net in the compact set X_i. □

Metrizability

The following is a result on ability for a topology to be given by a metric:

9.24 Theorem (Urysohn metrizability theorem). *A regular space with a countable basis is metrizable.*

The proof uses the Urysohn lemma and the metrizability of the Hibert cube. There are more general metrizability theorems, see [Mun00, p. 243].

Space filling curves

9.25 Theorem. *There is a continuous curve filling a rectangle on the plane. More concisely, there is a continuous map from the interval* $[0,1]$ *onto the square* $[0,1]^2$ *under the Euclidean topology.*

Such a curve is called a Peano curve. It could be constructed as a limit of an iteration of piecewise linear curves [Mun00, p. 271].

Part II Algebraic Topology

Briefly, Algebraic Topology associates algebraic objects, such as numbers, groups, vector spaces, modules, … to topological objects, then studies these algebraic objects in order to understand more about the topological objects.

This course aims to help students to grasp fundamental notions, methods, and results of beginning combinatorial and algebraic topology. The course focuses on discussing important and representative examples. Certain results which are fundamental but difficult to prove, such as equivalence among homologies, are assumed. The course benefits students interested in theoretical mathematics. Recently certain aspects of algebraic topology have been applied to computational sciences, so students interested in applied mathematics or computer science might also find this course beneficial.

Taking a prior course in General Topology is not required, however is strongly recommended.

There are several books presenting aspects of algebraic topology to undergraduate students, such as [Cro78], [Ams83], [J84]. The book [Mun00] has a part on algebraic topology but stops before homology. The book [Vas01] gives an overview of many aspects of both algebraic and differential topology.

Aiming at graduate students, the book [Hat01] can be consulted. Our choice of topics turns out to be similar to that in the book [Lee11], where more details and more advanced treatments can be found.

10 Structures on topological spaces

Topological manifold

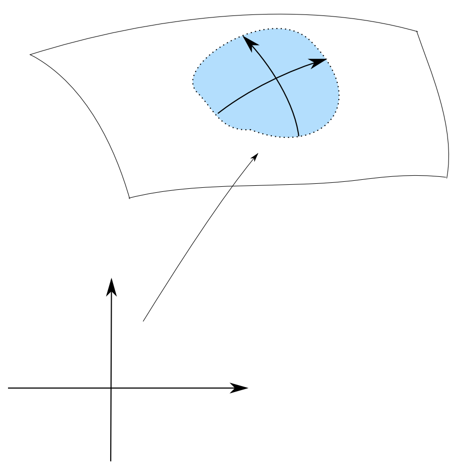
If we only stay around our small familiar neighborhood then we might not be able to recognize that the surface of the Earth is curved, and to us it is indistinguishable from a plane. When we begin to travel farther and higher, we can realize that the surface of the Earth is a sphere, not a plane. In mathematical language, a sphere and a plane are locally same but globally different.

Nowadays we frequently use maps of the surface of the Earth for our daily life. We look at a picture or a drawing (a rectangle on a plane) which provides a description (meaning a bijective correspondence which respects certain properties) of a region on the surface of the Earth. To get more information we often look at many such pictures of various regions, which can be overlapping. Previously such a set of pictures or drawing printed on paper providing coverage of a region on the surface of the Earth was often called an atlas.

We could say that taking lots of pictures of an object, a very popular activity nowadays, is the action of providing a description of the object by taking many maps of neighborhoods of the object to planes. This indeed corresponds to the mathematical the idea of manifold.

Another idea is that a manifold is a space which can be locally described by real numbers. The descriptions vary from one part to another part of the manifold, each such description is like a page of a map while the whole collection of the pages is an atlas providing a map of the whole manifold.

Definition. *A* **topological manifold** [1] *of dimension n is a topological space each point of which has a neighborhood homeomorphic to the Euclidean space $\mathbb{R}^n$.*



Briefly, ***a manifold is a space that is locally Euclidean***. [2]

[1] đa tạp tôpô

[2] Bernhard Riemann proposed the idea of manifold in 1854.

We can think of a manifold as a space that could be covered by a collection of open subsets each of which homeomorphic to $\mathbb{R}^n$.

Remark. In this part on Algebraic Topology we assume $\mathbb{R}^n$ has the Euclidean topology unless we mention otherwise.

The statement below is often convenient in practice:

Proposition. *A manifold of dimension n is a space such that each point has an open neighborhood homeomorphic to an open subset of $\mathbb{R}^n$.*

Proof. Let M be a manifold. Suppose that U is a neighborhood of x in M and $\phi : U \to \mathbb{R}^n$ is a homeomorphism. There is an open subset U' of M such that $x \in U' \subset U$. Since ϕ is a homeomorphism on U and U' is open in U, the set $\phi(U')$ is open in $\mathbb{R}^m$. Take a ball $B(\phi(x), r) \subset \phi(U')$. Since ϕ is continuous on U', the set $U'' = \phi^{-1}(B(\phi(x), r))$ contains x and is open in U' hence is open in M. The restriction $\phi|_{U''} : U'' \to B(\phi(x), r)$ is a homeomorphism. We have just shown that any point in the manifold has an open neighborhood homeomorphic to an open ball in $\mathbb{R}^n$.

For the reverse direction, suppose that M is a topological space, $x \in M$, U is an open neighborhood of x in M and $\phi : U \to V$ is homeomorphism where V is open in $\mathbb{R}^n$. Take a ball $B(\phi(x), r) \subset V$ and let $U' = \phi^{-1}(B(\phi(x), r)$ then U' contains x and is an open set in U hence is open in M. The restriction $\phi|_{U'} : U' \to B(\phi(x), r)$ is a homeomorphism. Recall that any open ball in $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^n$ (see 3.3), hence the ball $B(\phi(x), r)$ is homeomorphic to $\mathbb{R}^n$ via a homeomorphism ψ. Then $\psi \circ \phi|_{U'}$ is a homeomorphism from U' to $\mathbb{R}^n$. □

Remark. By Invariance of dimension (18.4), $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic unless $m = n$, therefore a manifold has a unique dimension.

Example. Any open subspace of $\mathbb{R}^n$ is a manifold of dimension n.

Example. Let $f : D \to \mathbb{R}$ be a continuous function where $D \subset \mathbb{R}^n$ is an open set, then the graph of f, the set $\{(x, f(x)) \mid x \in D\}$, as a subspace of $\mathbb{R}^{n+1}$, is homeomorphic to D, see 7.12, therefore is an n-dimensional manifold.

Thus manifolds generalizes curves and surfaces.

Example. The sphere S^n is an n-dimensional manifold. One way to show this is by covering S^n with two neighborhoods $S^n \setminus \{(0, \ldots, 0, 1)\}$ and $S^n \setminus \{(0, \ldots, 0, -1)\}$. Each of these neighborhoods is homeomorphic to $\mathbb{R}^n$ via stereographic projections. Another way is by covering S^n by hemispheres $\{(x_1, x_2, \ldots, x_{n+1}) \in S^n \mid x_i > 0\}$ and $\{(x_1, x_2, \ldots, x_{n+1}) \in S^n \mid x_i < 0\}$, $1 \leq i \leq n+1$. Each of these hemispheres is a graph, homeomorphic to an open n-dimensional unit ball.

Example. The torus is a two-dimensional manifold. Let us consider the torus as the quotient space of the square $[0,1]^2$ by identifying opposite edges. Each point has a neighborhood homeomorphic to an open disk, as can be seen easily in the following figure, though explicit description would be time consuming. We can also view the torus as a surface in $\mathbb{R}^3$, given by the equation $(\sqrt{x^2+y^2}-$

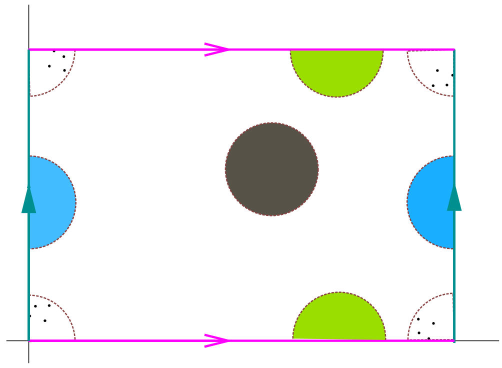


Figure 10.1: The sets with same colors are glued to form a neighborhood of a point on the torus. Each such neighborhood is homeomorphic to an open ball.

$a)^2+z^2=b^2$, see Fig. 8.10. As such it can be covered by the open subsets of $\mathbb{R}^3$ corresponding to $z>0$, $z<0$, $x^2+y^2<a^2$, $x^2+y^2>a^2$.

Remark. The interval $[0,1]$ is not a manifold, it is a "manifold with boundary", see section 24.

Simplicial complex

For an integer $n \geq 0$, an n-dimensional **simplex** [1] is a subspace of a Euclidean space $\mathbb{R}^m$, $m \geq n$ which is the set of all convex linear combinations of $(n+1)$ points in $\mathbb{R}^m$ which do not belong to any n-dimensional hyperplane. As a set it is given by $\{t_0v_0+t_1v_1+\cdots+t_nv_n \mid t_0,t_1,\ldots,t_n \in [0,1],\ t_0+t_1+\cdots+t_n = 1\}$ where $v_0,v_1,\ldots,v_n \in \mathbb{R}^m$ and $v_1-v_0, v_2-v_0,\ldots,v_n-v_0$ are n linearly independent vectors (it can be checked in Problem 10.18 that this condition does not depend on the order of the points). The points $v_0,v_1,\ldots,v_n$ are called the **vertices** of the simplex.

Example. A 0-dimensional simplex is just a point. A 1-dimensional simplex is a straight segment in $\mathbb{R}^m$, $m \geq 1$. A 2-dimensional simplex is a triangle in $\mathbb{R}^m$, $m \geq 2$. A 3-dimensional simplex is a tetrahedron in $\mathbb{R}^m$, $m \geq 3$.

In particular, the **standard$n$-dimensional simplex** [2] Δ_n is the convex linear combination of the $(n+1)$ vectors $(1,0,0,\ldots)$, $(0,1,0,0,\ldots)$, …, $(0,0,\ldots,0,1)$

[1] đơn hình
[2] đơn hình chuẩn

in $\mathbb{R}^{n+1}$. Thus

$$\Delta_n = \{(t_0, t_1, \ldots, t_n) \mid t_0, t_1, \ldots, t_n \in [0,1],\ t_0 + t_1 + \cdots + t_n = 1\}.$$

The convex linear combination of any subset of the set of vertices of a simplex is called a **face** of the simplex.

Example. For a 2-dimensional simplex (a triangle) its faces are the vertices, the edges, and the triangle itself.

Definition. *An n-dimensional* **simplicial complex** [1] *in $\mathbb{R}^m$ is a finite collection S of simplices in $\mathbb{R}^m$ of dimensions at most n and at least one simplex is of dimension n and such that:*

(a) any face of an element of S is an element of S,

(b) the intersection of any two elements of S is a common face.

The union of all elements of S is called its **underlying space**, denoted by $|S|$, a subspace of $\mathbb{R}^m$. A space which is the underlying space of a simplicial complex is also called a **polyhedron** [2].

Example. A 1-dimensional simplicial complex is often called a (combinatorial) graph in graph theory.

Triangulation

A **triangulation** [3] of a topological space X is a homeomorphism from the underlying space of a simplicial complex to X, the space X is then said to be **triangulated**.

For example, a triangulation of a surface is an expression of the surface as the union of finitely many triangles, with a requirement that two triangles are either disjoint, or have one common edge, or have one common vertex.

A simplicial complex is specified by a finite set of points, if a space can be triangulated then we can study that space combinatorially, using constructions and computations in finitely many steps.

Cell complex

For $n \geq 1$ a ***cell*** (ô) is an open ball in the Euclidean space $\mathbb{R}^n$. A 0-dimensional cell is a point (this is consistent with the general case with the convention that $\mathbb{R}^0 = \{0\}$).

---

[1] phức đơn hình
[2] đa diện
[3] phép phân chia tam giác

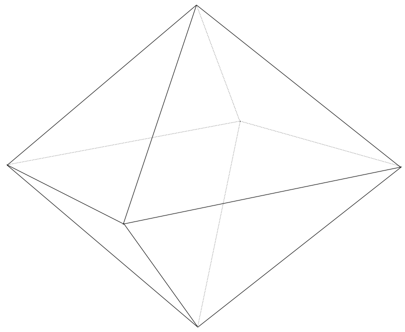


Figure 10.2: This is an octahedron, giving a triangulation of the 2-dimensional sphere.

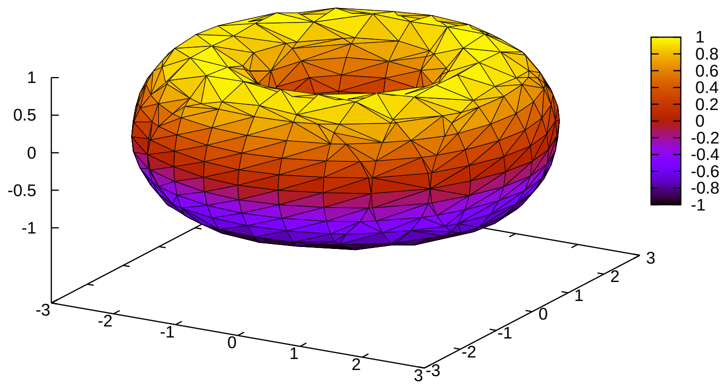


Figure 10.3: A triangulation of the torus.

Recall a familiar term that an n-dimensional **disk** is a closed ball in $\mathbb{R}^n$, in particular when $n = 0$ it is a point. The unit disk centered at the origin $B'(0,1)$ is denoted by D^n. Thus for $n \geq 1$ the boundary ∂D^n is the sphere S^{n-1} and the interior $\text{int}(D^n)$ is an n-cell.

By **attaching a cell**to a topological space X we mean taking a continuous function $f : \partial D^n \to X$ then forming the quotient space $(X \sqcup D^n)/(x \sim f(x),\ x \in \partial D^n)$ (for disjoint union see 7.24). Intuitively, we attach a cell to the space by gluing each point on the boundary of the disk to a point on the space in a certain way.

We can attach finitely many cells to X in the same manner. Precisely, attaching k n-cells to X means taking the quotient space $\left(X \sqcup \left(\bigsqcup_{i=1}^k D_i^n\right)\right)/(x \sim f_i(x),\ x \in \partial D_i^n, 1 \leq i \leq k)$ where $f_i : \partial D_i^n \to X$ is continuous, $1 \leq i \leq k$.

Definition. *A (finite) n-dimensional***cell complex** *(phức ô) or* **CW-complex** *X is a topological space with a structure as follows:*

(a) X^0 is a finite collection of 0-cells, with the discrete topology,

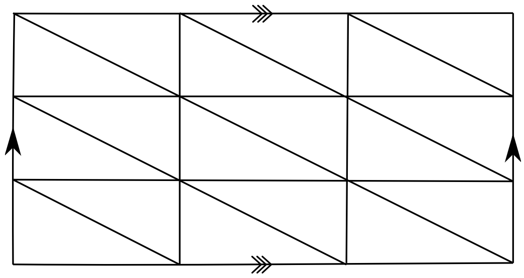


Figure 10.4: Description for a triangulation of the torus.

(b) for $1 \le i \le n \in \mathbb{Z}^+$, X^i is obtained by attaching finitely many i-cells to X^{i-1}, and X^n is homeomorphic to X.

Briefly, a cell complex is a topological space with an instruction for building it by attaching cells. The subspaces X^i are called the i-dimensional **skeleton** [1] of X.[2]

Example. A topological circle has a cell complex structure as a triangle with three 0-cells and three 1-cells. There is another cell complex structure with only one 0-cell and one 1-cell.

Example. The 2-dimensional sphere has a cell complex structure as a tetrahedron with four 0-cells, six 1-cells, and four 2-cells. There is another cell complex structure with only one 0-cell and one 2-cell, obtained by gluing the boundary of a 2-disk to a point.

Example. The torus, as we can see directly from its definition (Fig. 8.8), has a cell complex structure with one 0-cells, two 1-cells, and one 2-cells.

A simplicial complex gives rise to a cell complex:

Proposition. *Any triangulated space is a cell complex.*

Proof. Let X be a simplicial complex. Let X^i be the union of all simplices of X of dimensions at most i. Then X^{i+1} is the union of X^i with finitely many $(i+1)$-dimensional simplices. Let Δ^{i+1} be such an $(i+1)$-dimensional simplex. The i-dimensional faces of Δ^{i+1} are simplices of X, so the union of those faces, which is the boundary of Δ^{i+1}, belongs to X^i. There is a homeomorphism from an $(i+1)$-disk to Δ^{i+1} (see 10.19), bringing the boundary of the disk to the boundary of Δ^{i+1}. Thus including Δ^{i+1} in X implies attaching an $(i+1)$-cell to X^i. □

The example of the torus indicates that cell complexes may require less cells than simplicial complexes. On the other hand we loose the combinatorial setting, because we need to specify the attaching maps.

[1] khung

[2] The term CW-complex is more general than the term cell complex, can be used when there are infinitely many cells.

It is known that any compact manifold of dimension different from 4 has a cell complex structure. Whether that is true or not in dimension 4 is not known yet [Hat01, p. 529].

Euler characteristic

Definition. *The* **Euler characteristic** [1] *of a cell complex is defined to be the alternating sum of the number of cells in each dimension of that cell complex. Namely, let X be an n-dimensional cell complex and let c_i, $0 \le i \le n$, be the number of i-dimensional cells of X, then the Euler characteristic of X is*

$$\chi(X) = \sum_{i=0}^{n} (-1)^i c_i.$$

Example. For a triangulated surface, its Euler characteristic is the number v of vertices minus the number e of edges plus the number f of triangles (faces):

$$\chi(S) = v - e + f.$$

10.5 Theorem. *The Euler characteristic of homeomorphic cell complexes are equal.*

The theorem can be proved by expressing the Euler characteristic of a cell complex through cellular homology, and by using the topological invariance of cellular homology, see Section 18, [Hat01, p. 146].

In particular, two cell complex structures on the same topological space have same Euler characteristics. Thus ***the Euler characteristic is a topological invariant of cell complexes***.

Example. The Euler characteristics of a tetrahedron, the octahedron in Fig. 10.2, the dodecahedron in 10.6, are 2. These spaces are all homeomorphic to the sphere S^2. We say that the Euler characteristics of of the sphere S^2 is 2, $\chi(S^2) = 2$.

Since any convex polyhedron is homeomorphic to the sphere S^2, see 17.22, we obtain the famous formula of Leonhard Euler:

Corollary (Euler formula). *For any convex polyhedron, $v - e + f = 2$.*

Example.

[1] đặc trưng, đặc số Euler

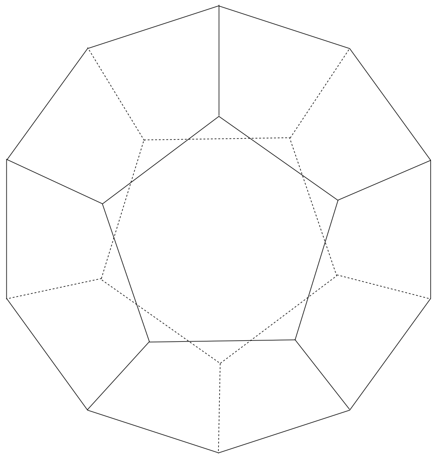


Figure 10.6: The Euler characteristic of the dodecahedron is 2.

Figure 10.7: This is a cell complex structure of a surface homeomorphic to the two-dimensional sphere, drawn by the software Blender, which reports that it consists of 40 vertices, 109 edges, and 71 faces, giving the Euler characteristic as 2.

Example. From the triangulation of the torus in Fig. 10.4, we get $\chi(T^2) = 0$.

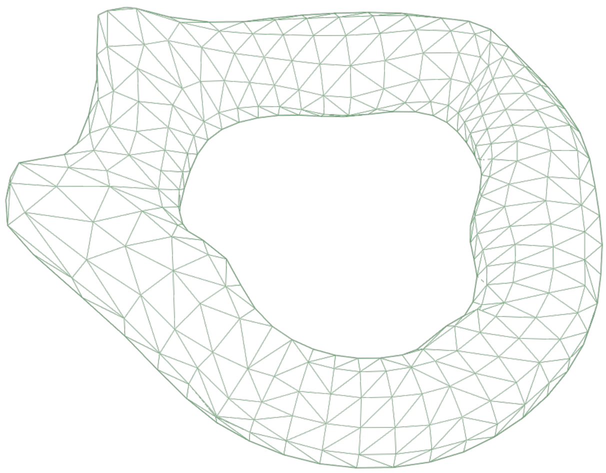


Figure 10.8: This is a triangulation of a surface homeomorphic to the torus, drawn by the software Blender, which reports that it consists of 576 vertices, 1728 edges, and 1152 faces, giving the Euler characteristic as 0.

Example. From the triangulation of the projective plane in Fig. 10.9, we get $\chi(\mathbb{R}P^2) = 1$.

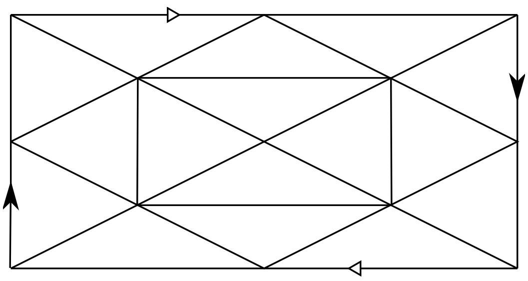


Figure 10.9: Description for a triangulation of the projective plane.

Example. The sphere, the torus, and the projective plane have different Euler characteristics, so they are not homeomorphic to each other.

Problems

10.10. Show that if two spaces are homeomorphic and one space is an n-dimensional manifold then the other is also an n-dimensional manifold.

10.11. Show that an open subspace of a manifold is a manifold.

10.12. Show that if X and Y are manifolds then $X \times Y$ is also a manifold.

10.13. Show that a connected manifold must be path-connected. Thus for manifold connectedness and path-connectedness are same.

10.14. Show that $\mathbb{R}P^n$ is an n-dimensional topological manifold.

10.15. Show that a manifold is a locally compact space.

10.16. Show that the Mobius band, without its boundary circle, is a manifold.

10.17. Give examples of topological spaces which are not manifolds. Discuss necessary conditions for a topological space to be a manifold.

10.18. Given a set of n points in $\mathbb{R}^n$, list the points as $v_0, v_1, \ldots, v_n$. Check that the condition that $v_1 - v_0, v_2 - v_0, \ldots, v_n - v_0$ are linearly independent vectors does not depend on the choice of orders for the points.

10.19. ✓ Show that any n-dimensional simplex is homeomorphic to an n-dimensional disk.

10.20. Give a triangulation for the cylinder, find a simpler cell complex structure for it, and compute its Euler characteristic.

10.21. Give a triangulation for the Mobius band, find a simpler cell complex structure for it, and compute its Euler characteristic.

10.22. Give a triangulation for the Klein bottle, find a simpler cell complex structure for it, and compute its Euler characteristic.

10.23. Give examples of spaces which are not homeomorphic but have same Euler characteristics.

10.24. Draw a cell complex structure on the torus with two holes.

10.25. Find a cell complex structure on $\mathbb{R}\mathrm{P}^n$.

10.26. Continuing 8.55, in combinatorics, given a combinatorial graph $G = (V, E, f)$, a path from vertex v_1 to vertex v_n is a sequence of edges $(e_1, e_2, \ldots, e_n)$ such that for $1 \le i \le n-1$ we have $f(e_i) = \{v_i, v_{i+1}\}$. Two vertices u and v are said to be connected if there is a path going from u to v. The combinatorial graph is said to be connected if there is a path between any two vertices.

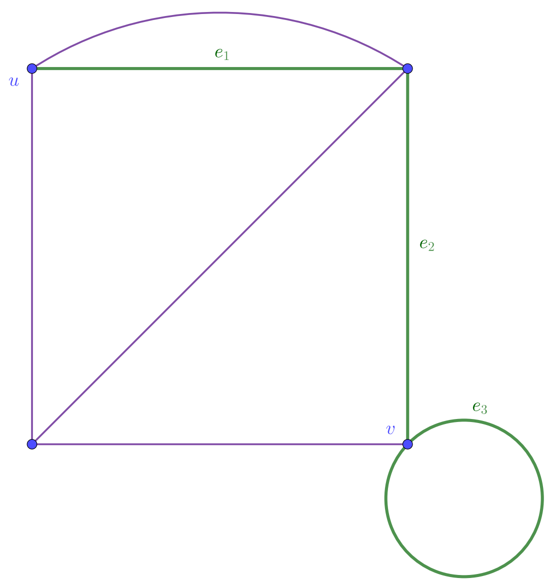


Figure 10.27: A drawing of a combinatorial graph and a path (e_1, e_2, e_3) from u to v.

The problem "Seven bridges of Konigsberg" can be studied as a problem about paths in the combinatorial graph in Fig 8.56.

We see that for combinatorial graph the notion of connectedness is introduced without topology. It is not clear whether in certain setting this notion of combinatorial graph connectedness is related to the notion of connectedness in topology.

In 8.55 we considered the following approach for topological graphs. Given a combinatorial graph $G = (V, E, f)$, let V have the discrete topology. For each $e \in E$, let I_e be the interval $[0, 1]$ with the Euclidean topology. Form the disjoint union

$$V \sqcup \left(\bigsqcup_{e \in E} I_e \right).$$

Take the quotient space by gluing $\partial I_e = \{0, 1\}$ with the vertices $f(e)$ of e. This is our **topological graph**:

$$\tilde{G} = \left[V \sqcup \left(\bigsqcup_{e \in E} I_e \right) \right] / \left[\partial I_e \sim f(e) \right].$$

Check that:

(a) The topological graph $\tilde{G}$ is a cell complex, assuming V and E are finite sets.

(b) If the combinatorial graph G is connected then the topological graph $\tilde{G}$ is path-connected.

(c) The topological graph $\tilde{G}$ is locally path-connected, hence it is path-connected if and only if it is connected.

For more on this topic, and related topics such as graphs of groups, see [Hat01, p. 83].

10.28 (Platonic solids). In this problem, using a term in classical geometry, by a convex polyhedron (đa diện lồi) in $\mathbb{R}^3$ we mean the union of a finite number of polygons (đa giác) in $\mathbb{R}^3$ such that any two such polygons intersect in either an empty set, or a common vertex, or a common edge, that each edge belong to exactly two polygons, and that it is the boundary of a convex compact subset of $\mathbb{R}^3$ with non-empty interior, called the corresponding solid. Notice that this is like the underlying space of a 2-dimensional simplicial complex except that the 2-dimensional faces are allowed to be polygons instead of only triangles. A regular convex polyhedron (đa diện lồi đều) is a convex polyhedron whose faces are the same regular polygons (đa giác đều) and each vertex belongs to the same number of faces.

Consider a regular convex polyhedron. Let p be the number of regular polygons at each vertex, and let q be the number of vertices of the regular polygon.

(a) Counting the number of vertices v from the number of faces f, show that $pv = qf$.

(b) Counting the number of edges e from the number of faces, show that $2e = qf$.

(c) It is known that a convex polyhedron is homeomorphic to a sphere (Problem 17.22). Using Euler characteristic, deduce that $f = \frac{4p}{2(p+q)-pq}$, and hence $2(p + q) - pq > 0$.

(d) Deduce that $3 \leq p < \frac{2q}{q-2}$.

(e) Deduce that there are only 5 possibilities for (p, q): $(3, 3)$, $(4, 3)$, $(5, 3)$, $(3, 4)$, $(3, 5)$.

It was known since ancient time that there exists those regular convex polyhedron: $(3,3)$ gives the regular tetrahedron, $(4,3)$ gives the regular octahedron (Figure 10.2), $(5,3)$ gives the regular icosahedron, $(3,4)$ gives the cube, $(3,5)$ gives the regular dodecahedron (Figure 10.6). The corresponding solids are called the Platonic solids.

11 Classification of compact surfaces

In this section by a **surface** [1] we mean a two-dimensional topological manifold (without boundary).

Given a polygon on the plane and suppose that the edges of the polygon are labeled and oriented. Choose one edge as the initial one then follow the edge of the polygon in a predetermined direction (the boundary of the polygon is homeomorphic to a circle). If an edge a is met in the opposite direction then write it down as a^{-1}. In this way we associate each polygon with a word. We also write aa as a^2.

In the reverse direction, a word gives a polygon with labeled and oriented edges. If a label appears more than once, then the edges with this label are identified in the orientations assigned to the edges. Let us consider two words **equivalent** if they give rise to homeomorphic spaces. For examples, changing labels and cyclic permutations are equivalence operations on words.

11.1 Theorem. *A connected compact surface is homeomorphic to the space obtained by identifying the edges of a polygon in one of the following ways:*

(a) aa^{-1},

(b) $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1}$,

(c) $c_1^2c_2^2\cdots c_g^2$.

The surface of the first type is clearly a sphere. We denote by T_g the surface of the second type, and by M_g the surface of the third type. The number g is called the **genus** [2] of the surface.

Thus any compact surface can be obtained by gluing the boundary of a disk.

Proof of 11.1

We shall use the following fact: any compact surface can be triangulated (a proof is available in [Moi77]).

Take a triangulation X of the compact connected surface S. Let T be the set of triangles in X. Label the edges of the triangles and mark an orientation for each edge. Each edge will appear twice, on two different triangles.

We now build a new space P out of these triangles. Let P_1 be any element of T, which is homeomorphic to a disk. Inductively, suppose we have built a space P_i from a subset T_i of T, where P_i is homeomorphic to a disk and its boundary consists of labeled and directed edges. If $T_i \neq T$ then there exists an

[1] mặt

[2] giống, chi

element $\Delta \in T \setminus T_i$ such that Δ has at least one edge with the same label α as an edge on the boundary of P_i, otherwise $|X|$ could not be connected. Glue Δ to P_i via this edge to get P_{i+1}, in other words $P_{i+1} = (P_i \sqcup \Delta)/\alpha \sim \alpha$. Since two disks glued along a common arc on the boundaries is still a disk, P_{i+1} is homeomorphic to a disk. Let $T_{i+1} = T_i \cup \{\Delta\}$. When this process stops we get a space P homeomorphic to a disk, whose boundary consists of labeled, directed edges. If these edges are identified following the instruction by the labels and the directions, we get a space homeomorphic to the original surface S. The polygon P is called a **fundamental polygon** of the surface S. See an example in Fig. 11.2.

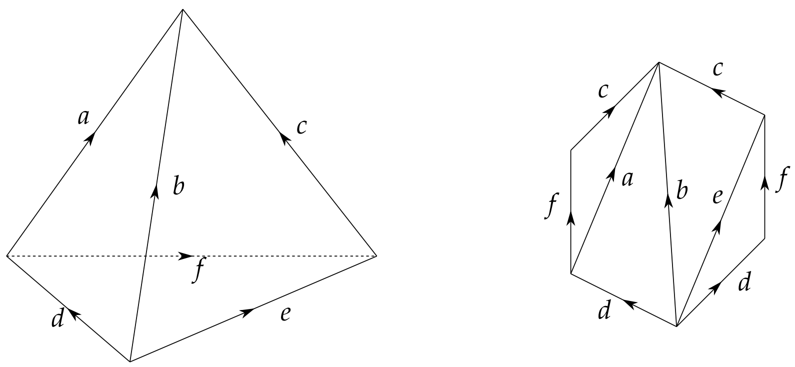


Figure 11.2: A triangulation of the sphere and an associated fundamental polygon.

Theorem 11.1 is a direct consequence of the following:

11.3 Proposition. *An associated word to a connected compact surface is equivalent to a word of the forms:*

(a) aa^{-1},

(b) $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}\cdots a_gb_ga_g^{-1}b_g^{-1}$,

(c) $c_1^2c_2^2\cdots c_g^2$.

We will prove 11.3 through a sequence of lemmas. The technique is cut-and-paste topology in Section 8.

Let w a the word associated to a fundamental polygon of a compact connected surface. If w_1 and w_2 are two equivalent words then we write $w_1 \sim w_2$.

11.4 Lemma. $aa^{-1}\alpha \sim \alpha$, *if* $\alpha \neq \emptyset$.

Proof. see Fig. 11.5.

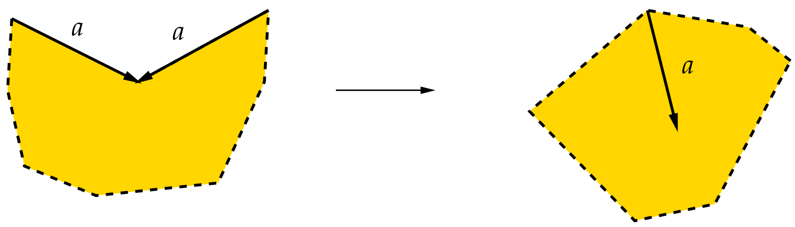


Figure 11.5: Lemma 11.4.

□

11.6 Lemma. *The word w is equivalent to a word whose all of the vertices of the associated polygon is identified to a single point on the associated surface (w is said to be "reduced").*

Proof. When we do the operation in Fig. 11.7, the number of P vertices is decreased. When there is only one P vertex left, we arrive at the situation in Lemma 11.4.

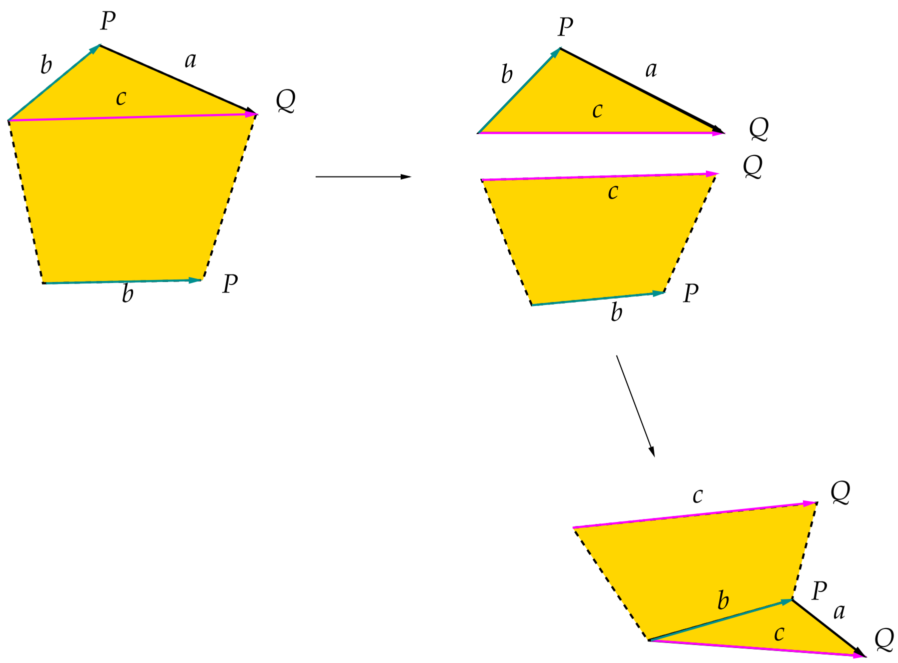


Figure 11.7: Lemma 11.6.

□

11.8 Lemma. $a\alpha a\beta \sim aa\alpha\beta^{-1}$*, where α and β are some words.*

Proof. see Fig. 11.9.

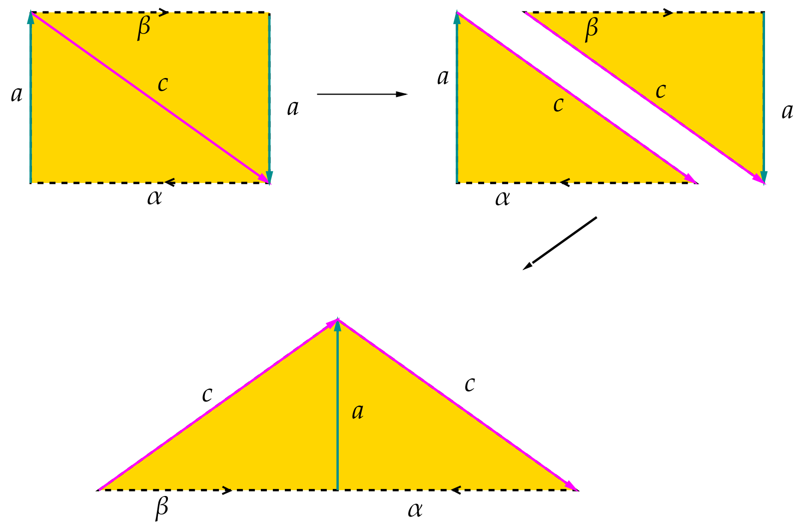


Figure 11.9: Lemma 11.8.

□

11.10 Lemma. *Suppose that w is reduced. If $w = a\alpha a^{-1}\beta$ where $\alpha \neq \emptyset$ then $\exists b \in \alpha$ such that $b \in \beta$ or $b^{-1} \in \beta$.*

Proof. If all labels in α appear in pairs then the vertices in the part of the polygon associated to α are identified only with themselves, and are not identified with a vertex outside of that part. This contradicts the assumption that w is reduced. □

11.11 Lemma. $a\alpha b\beta a^{-1}\gamma b^{-1}\delta \sim aba^{-1}b^{-1}\alpha\delta\beta\gamma$.

Proof. see Fig. 11.12.

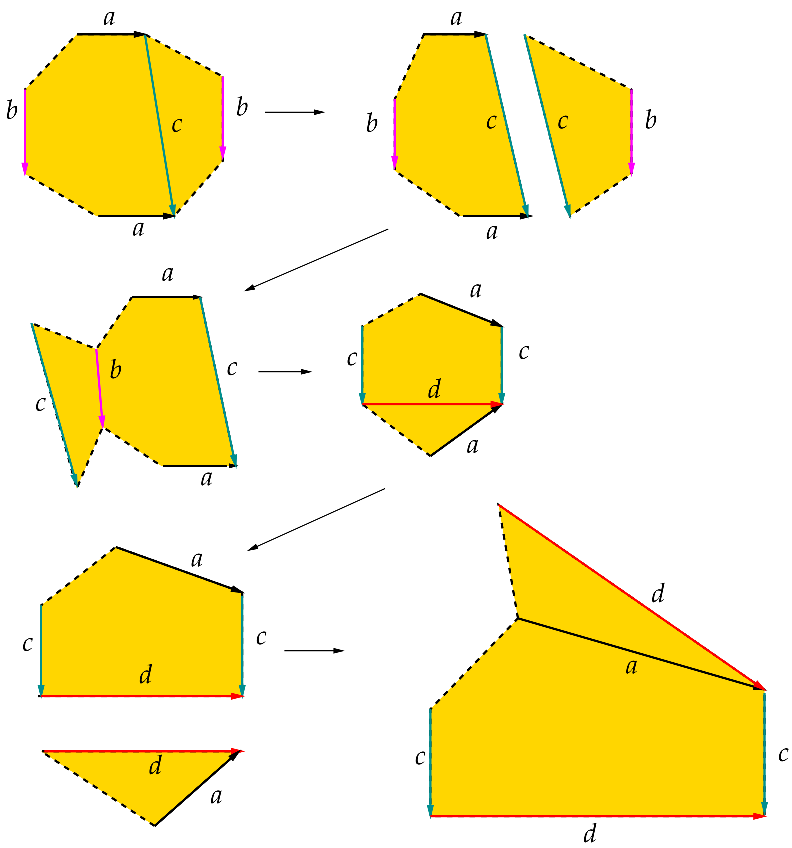


Figure 11.12: Lemma 11.11.

□

11.13 Lemma. $aba^{-1}b^{-1}\alpha c^2\beta \sim a^2b^2\alpha c^2\beta$.

Proof. Do the operation in figure 11.14, after that we are in a situation where we can apply lemma 11.8 three times.

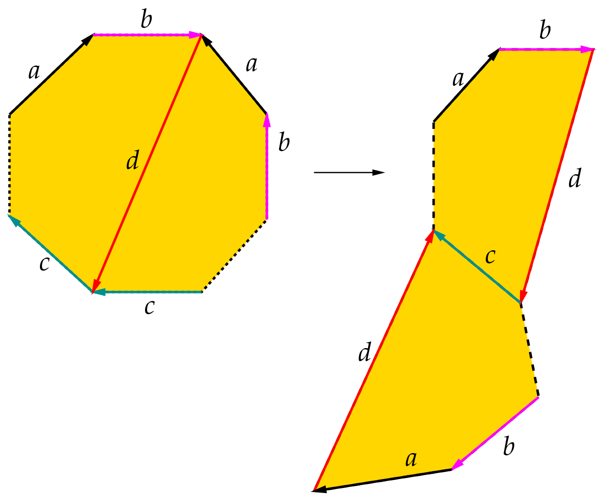


Figure 11.14: Lemma 11.13.

□

Proof of 11.3. The proof follows the following steps.

1. Bring w to reduced form by using 11.6 finitely many times.

2. If w has the form $aa^{-1}\alpha$ then go to 2.1, if not go to 3.

2.1. If w has the form aa^{-1} then stop, if not go to 2.2.

2.2. w has the form $aa^{-1}\alpha$ where $\alpha \neq \emptyset$. Repeatedly apply 11.4 finitely many times, deleting pairs of the form aa^{-1} in w until no such pair is left or w has the form aa^{-1}. If no such pair is left go to 3.

3. w does not have the form $\alpha aa^{-1}\beta$. If we apply 11.8 then a pair of of the form $-a\alpha a-$ with $\alpha \neq \emptyset$ could become a pair of the form $-a - a^{-1}-$, but a pair of the form $-aa-$ will not be changed. Therefore 11.8 could be used finitely many times until there is no pair $-a\alpha a-$ with $\alpha \neq \emptyset$ left. Notice from the proof of 11.8 that this step will not undo the steps before it.

4. If there is no pair of the form $-a\alpha a^{-1}-$ where $\alpha \neq \emptyset$, then stop: w has the form $a_1^2 a_2^2 \cdots a_g^2$.

5. w has the form $-a\alpha a^{-1}-$ where $\alpha \neq \emptyset$. By 11.10 w must has the form $-a - b - a^{-1} - b^{-1}-$, since after Step 3 there could be no $-b - a^{-1} - b-$.

6. Apply 11.11 finitely many times until w no longer has the form $-a\alpha b\beta a^{-1}\gamma b^{-1}-$ where at least one of α, β, or γ is non-empty. If w is not of the form $-aa-$ then stop: w has the form $a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1}$.

7. w has the form $-aba^{-1}b^{-1} - cc-$. Use 11.13 finitely many times to transform w to the form $a_1^2 a_2^2 \cdots a_g^2$.

□

Example. What is the topological space in Fig. 11.15?

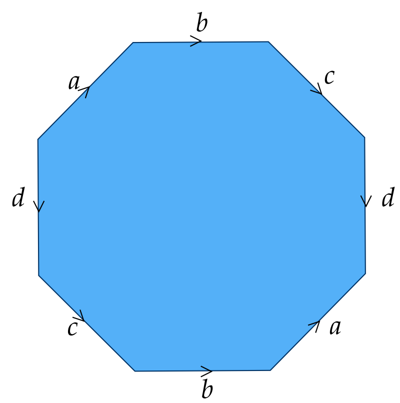


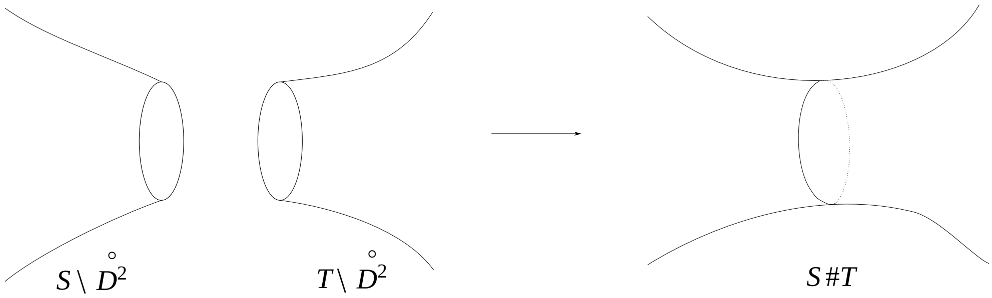
Figure 11.15

Applying lemma 11.11 we get $abcda^{-1}b^{-1}c^{-1}d^{-1} \sim aba^{-1}b^{-1}cdc^{-1}d^{-1}$. Thus this space is homeomorphic to the genus 2 torus.

For more on this topic one can read [Mas91, ch. 1].

Connected sum

Let S and T be two surfaces. From each surface deletes an open disk, then glue the two boundary circles. The resulting surface is called the **connected sum** [1] of the two surfaces, denoted by $S\#T$.



It is known that the connected sum does not depend on the choices of the disks.

Example. If S is any surface then $S\#S^2 = S$.

Theorem (classification of compact surfaces). *A connected compact surface is homeomorphic to either the sphere, or a connected sum of tori [2], or a connected sum of projective planes.*

The sphere and the surfaces T_g are **orientable**[3] surface , while the surfaces M_g are **non-orientable** [4] surfaces. We will not give a precise definition of orientability here (compare 25).

Problems

11.17. Show that $\mathbb{RP}^2\#\mathbb{RP}^2 = K$, in other words, gluing two Mobius bands along their boundaries gives the Klein bottle. see Fig. 8.22.

11.18. Show that $T^2\#\mathbb{RP}^2 = K\#\mathbb{RP}^2$, where K is the Klein bottle.

[1] tổng liên thông
[2] The plural form of the word torus is tori.
[3] định hướng được
[4] không định hướng được

Figure 11.16: Orientable surfaces: $S^2, T_1, T_2, \ldots$

11.19. What is the topological space in Fig. 11.20?

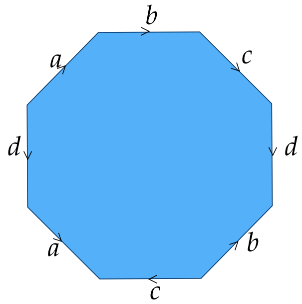


Figure 11.20

11.21. (a) Show that $T_g \# T_h = T_{g+h}$.

(b) Show that $M_g \# M_h = M_{g+h}$.

(c) $M_g \# T_h = ?$

11.22. Show that for any two compact surfaces S_1 and S_2 we have $\chi(S_1 \# S_2) = \chi(S_1) + \chi(S_2) - 2$.

11.23. Compute the Euler characteristics of all connected compact surfaces. Deduce that the orientable surfaces S^2 and T_g, for different g, are distinct, meaning not homeomorphic to each other. Similarly the non-orientable surfaces M_g are all distinct.

11.24. From 11.1, describe a cell complex structure on any compact surface. From that compute the Euler characteristics.

11.25. Let S be a triangulated compact connected surface with v vertices, e edges, and f triangles. Prove that:

(a) $2e = 3f$. The number of triangles is even.

(b) $v \leq f$.

(c) $e \leq \frac{1}{2}v(v-1)$.

(d) $v \geq \left\lfloor \frac{1}{2}\left(7 + \sqrt{49 - 24\chi(S)}\right) \right\rfloor = H(S)$. The integer $H(S)$, called the Heawood number, gives the minimal number of colors needed to color a map on the surface S, except in the case of the Klein bottle. When S is the sphere this is The four colors problem [MT01, p. 230].

(e) Show that a triangulation of the torus needs at least 14 triangles. Indeed 14 is the minimal number: there are triangulations of the torus with exactly 14 triangles, see e.g. [MT01, p. 142].

12 Homotopy

Homotopy of maps

Definition. *Let X and Y be topological spaces and $f, g : X \to Y$ be continuous. We say that f and g are* **homotopic** [1] *if there is a continuous map*

$$\begin{aligned} F : X \times [0,1] &\to Y \\ (x,t) &\mapsto F(x,t) \end{aligned}$$

such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$ for all $x \in X$. The map F is called a **homotopy**[2] *from f to g.*

We can think of t as a time parameter and F as a continuous process in time that starts with f and ends with g. To suggest this view $F(x,t)$ is often written as $F_t(x)$. So $F_0 = f$ and $F_1 = g$.

Here we are looking at $[0,1] \subset \mathbb{R}$ with the Euclidean topology and $X \times [0,1]$ with the product topology.

12.1 Proposition. *Homotopic relation on the set of continuous maps between two given topological spaces is an equivalence relation.*

Proof. If $f : X \to Y$ is continuous then f is homotopic to itself via the map $F_t = f, \forall t \in [0,1]$. If f is homotopic to g via F then g is homotopic to f via $G(x,t) = F(x, 1-t)$. We can think of G is an inverse process to H. If f is homotopic to g via F and g is homotopic to h via G then f is homotopic to h via

$$H_t = \begin{cases} F_{2t}, & 0 \le t \le \frac{1}{2} \\ G_{2t-1}, & \frac{1}{2} \le t \le 1. \end{cases}$$

Thus we obtain H by following F at twice the speed, then continuing with a copy of G also at twice the speed. To check the continuity of H it is better to write it as

$$H(x,t) = \begin{cases} F(x,2t), & (x,t) \in X \times [0, \frac{1}{2}], \\ G(x, 2t-1), & (x,t) \in X \times [\frac{1}{2}, 1], \end{cases}$$

then use 3.29. □

In this section, and whenever it is clear from the context, we indicate this relation by the usual notation for equivalence relation $\sim$.

12.2 Lemma. *If $f, g : X \to Y$ and $f \sim g$ then $f \circ h \sim g \circ h$ for any $h : X' \to X$, and $k \circ f \sim k \circ g$ for any $k : Y \to Y'$.*

---

[1] đồng luân

[2] phép đồng luân

Proof. Let F be a homotopy from f to g then $G_t = F_t \circ h$ is a homotopy from $f \circ h$ to $g \circ h$. Rewriting $G(x', t) = F(h(x'), t)$ we see that G is continuous. □

Homotopic spaces

Definition. *Two topological spaces X and Y are* **homotopic** *if there are continuous maps $f : X \to Y$ and $g : Y \to X$ such that $g \circ f$ is homotopic to Id_X and $f \circ g$ is homotopic to Id_Y. Each of the maps f and g is called a* **homotopy equivalence**.

Immediately we have:

Proposition (homeomorphic ⇒ homotopic). *Homeomorphic spaces are homotopic.*

Proof. If f is a homeomorphism then we can take $g = f^{-1}$. □

Proposition. *Homotopic relation on the set of all topological spaces is an equivalence relation.*

Proof. Any space is homotopic to itself. From the definition it is clear that this relation is symmetric. Now we check that it is transitive. Suppose that X is homotopic to Y via $f : X \to Y$, $g : Y \to X$ and suppose that Y is homotopic to Z via $h : Y \to Z$, $k : Z \to Y$. By 12.2:

$$(g \circ k) \circ (h \circ f) = (g \circ (k \circ h)) \circ f \sim (g \circ \mathrm{Id}_Y) \circ f = g \circ f \sim \mathrm{Id}_X.$$

Similarly

$$(h \circ f) \circ (g \circ k) = (h \circ (f \circ g)) \circ k \sim (h \circ \mathrm{Id}_Y) \circ k = h \circ k \sim \mathrm{Id}_Z.$$

Thus X is homotopic to Z. □

Deformation retraction

Let X be a space, and let A be a subspace of X. We say that A is a **retract** [1] of X if there is a continuous map $r : X \to A$ such that $r|_A = \mathrm{id}_A$, called a **retraction** [2] from X to A. In other words A is a retract of X if the identity map id_A can be extended to X.

A **deformation retraction** [3] from X to A is a homotopy F that starts with id_X, ends with a retraction from X to A, and fixes A throughout, i.e., $F_0 = \mathrm{id}_X$, $F_1(X) = A$, and $F_t|_A = \mathrm{id}_A$, $\forall t \in [0, 1]$. If there is such a deformation retraction we say that A is a **deformation retract** [4] of X.

[1] rút
[2] phép rút
[3] phép co rút
[4] co rút

In such a deformation retraction each point $x \in X \setminus A$ "moves" along the path $F_t(x)$ to a point in A, while every point of A is fixed.

Example. A subset A of a normed space is called a **star-shaped** region if there is a point $x_0 \in A$ such that for any $x \in A$ the straight segment from x to x_0 is contained in A. For example, any convex subset of the normed space is a star-shaped region. The map $F_t = (1-t)x + tx_0$ is a deformation retraction from A to x_0, so a star-shaped region has a deformation retraction to a point.

Example. A normed space minus a point has a deformation retraction to a sphere. For example, a normed space minus the origin has a deformation retraction $F_t(x) = (1-t)x + t\frac{x}{||x||}$ to the unit sphere at the origin.

Example. An annulus $S^1 \times [0,1]$ has a deformation retraction to one of its circle boundary $S^1 \times \{0\}$.

Proposition. *If a space X has a deformation retraction to a subspace A then X is homotopic to A.*

Proof. Suppose that F_t is a deformation retraction from X to A. Consider $F_1 : X \to A$ and the inclusion map $g : A \to X$, $g(x) = x$. Then id_X is homotopic to $g \circ F_1$ via F_t, while $F_1 \circ g = \mathrm{id}_A$. □

Example. The circle, the annulus, and the Mobius band are homotopic to each other, although they are not homeomorphic to each other.

A space which is homotopic to a space containing only one point is said to be **contractible** [1]. It can be check immediately that this means there is a homotopy from the identity map to a constant map. Thus a space X is contractible if and only if there exists $x_0 \in X$ and a continuous map $F : X \times [0,1] \to X$ such that $F_0 = \mathrm{id}_X$ and $F_1 = x_0$.

Corollary. *A space which has a deformation retraction to a point is contractible.*

The converse is not true, since the point x_0 is not required to be fixed during the homotopy of a contractible space, see [Hat01, p. 18].

Example. Any star-shaped region is contractible.

Problems

12.3. Show in detail that the Mobius band has a deformation retraction to a circle.

12.4. Show that contractible spaces are path-connected.

12.5. Let X be a topological space and let Y be a retract of X. Show that any continuous map from Y to a topological space Z can be extended to X.

[1] thắt được

12.6. Let X be a topological space and let Y be a retract of X. Show that if every continuous map from X to itself has a fixed point then every continuous map from Y to itself has a fixed point.

12.7. Show that a retract of a Hausdorff space must be a closed subspace. Deduce that in a normed space, there exists a retraction from the space the the closed unit ball, but there is no retraction from the space to the open unit ball.

12.8. Show that if B is contractible then $A \times B$ is homotopic to A.

12.9. Show that if a space X is contractible then all continuous maps from a space Y to X are homotopic.

12.10. Let X be a topological space and let $f : X \to S^n$, $n \geq 1$, be a continuous map which is not surjective. Show that f is homotopic to a constant map.

12.11. Show that if $f : S^n \to S^n$ is not homotopic to the identity map then there is $x \in S^n$ such that $f(x) = -x$.

12.12 (isotopy). An isotopy is a homotopy by homeomorphisms. Namely two homeomorphisms f and g from X to Y are ***isotopic*** if there is a continuous map

$$\begin{aligned} F : X \times [0,1] &\to Y \\ (x,t) &\mapsto F(x,t) \end{aligned}$$

such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$ for all $x \in X$, and for each $t \in [0,1]$ the map $F(\cdot, t)$ is a homeomorphism. Check that being isotopic is an equivalence relation.

12.13 (Alexander trick). * Any homeomorphism of the disk fixing the boundary is isotopic to the identity.

This result can be proved in the following way. Suppose that f is homeomorphism from D^n onto D^n which is the identity on ∂D^n. Let $F : D^n \times [0,1] \to D^n$,

$$F(x,t) = \begin{cases} x, & \|x\| \geq t, \\ tf\left(\frac{x}{t}\right), & \|x\| < t. \end{cases}$$

At each $t > 0$ the map $F(\cdot, t)$ replicates f on a smaller ball $B'(0,t)$, so as $t \to 0$ gradually $F(\cdot, t)$ becomes the identity map. Check that F is indeed an isotopy from f to the identity.

12.14. Classify the alphabetical characters according to homotopy types, that is, which of the characters are homotopic to each other as subspaces of the Euclidean plane? Do the same for the Vietnamese alphabetical characters. Note that the result depends on the font you use. This is a continuation of the classification up to homeomorphisms at 4.32.

13 The fundamental group

A **path** [1] in a space X is a continuous map α from the Euclidean interval $[0,1]$ to X. The point $\alpha(0)$ is called the initial end point, and $\alpha(1)$ is called the final end point. In this section for simplicity of presentation we assume the domain of a path is the Euclidean interval $[0,1]$ instead of any Euclidean closed interval as before.

A **loop** [2] or a **closed path** [3] based at a point $a \in X$ is a path whose initial point and end point are both a. In other words it is a continuous map $\alpha : [0,1] \to X$ such that $\alpha(0) = \alpha(1) = a$. The **constant loop** at a is the loop $\alpha(t) = a$ for all $t \in [0,1]$.

If $\alpha(t)$, $0 \le t \le 1$ is a path from a to b then we write as α^{-1} the path given by $\alpha^{-1}(t) = \alpha(1-t)$, called the **inverse path** of α, going from b to a.

If α is a path from a to b, and β is a path from b to c, then we define the **composition** [4] of α with β, written $\alpha \cdot \beta$, to be the path

$$\alpha \cdot \beta(t) = \begin{cases} \alpha(2t), & 0 \le t \le \frac{1}{2} \\ \beta(2t-1), & \frac{1}{2} \le t \le 1. \end{cases}$$

Notice that $\alpha \cdot \beta$ is continuous by 3.29.

Homotopy of paths

Definition. *Let α and β be two paths from a to b in X. A* **path-homotopy** [5] *from α to β is a continuous map $F : [0,1] \times [0,1] \to X$, $F(s,t) = F_t(s)$, such that $F_0 = \alpha$, $F_1 = \beta$, and for each t the path F_t goes from a to b, i.e. $F_t(0) = a$, $F_t(1) = b$.*

If there is a path-homotopy from α to β we say that α is ***path-homotopic*** (đồng luân đường) to β.

Remark. A homotopy of path is a homotopy of maps defined on $[0,1]$, ***with the further requirement that the homotopy fixes the initial point and the terminal point***. To emphasize this we have used the word path-homotopy, but some sources (e.g. [Hat01, p. 25]) simply use the term homotopy, taking this further requirement implicitly.

13.2 Example. In a convex subset of a normed space any two paths α and β with the same initial points and end points are homotopic, via the homotopy $(1-t)\alpha + t\beta$.

[1] đường đi
[2] vòng
[3] đường đi kín
[4] hợp
[5] phép đồng luân đường

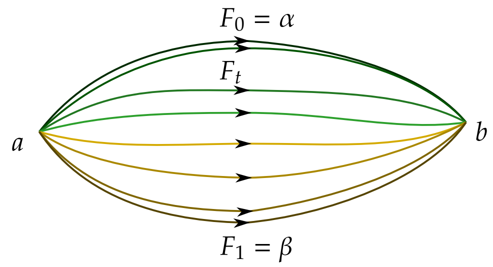


Figure 13.1: We can think of a path-homotopy from α to β as a way to continuously brings α to β, keeping the endpoints fixed.

Proposition. *Path-homotopic relation on the set of all paths from a to b is an equivalence relation.*

The proof is the same as the proof of 12.1.

13.3 Lemma. *If α is path-homotopic to α_1 and β is path-homotopic to β_1 then $\alpha \cdot \beta$ is path-homotopic to $\alpha_1 \cdot \beta_1$.*

Proof. Let F be a path-homotopy from α to β and let G be a path-homotopy from β to γ. Let $H_t = F_t \cdot G_t$, that is:

$$H(s,t) = \begin{cases} F(2s,t), & 0 \le s \le \frac{1}{2},\ 0 \le t \le 1 \\ G(2s-1,t), & \frac{1}{2} \le s \le 1,\ 0 \le t \le 1. \end{cases}$$

Then H is continuous by 3.29 and is a path-homotopy from $\alpha \cdot \beta$ to $\alpha_1 \cdot \beta_1$. □

13.4 Lemma. *If α is a path from a to b then $\alpha \cdot \alpha^{-1}$ is path-homotopic to the constant loop at a.*

Proof. A homotopy F from $\alpha \cdot \alpha^{-1}$ to the constant loop at a can be described as follows. At a fixed t, the loop F_t starts at time 0 at a, goes along α but at twice the speed of α, until time $\frac{1}{2} - \frac{t}{2}$, stays there until time $\frac{1}{2} + \frac{t}{2}$, then catches the inverse path α^{-1} at twice its speed to come back to a. More precisely,

$$F(s,t) = \begin{cases} \alpha(2s), & 0 \le s \le \frac{1}{2} - \frac{t}{2} \\ \alpha(1-t), & \frac{1}{2} - \frac{t}{2} \le s \le \frac{1}{2} + \frac{t}{2} \\ \alpha^{-1}(2s-1) = \alpha(2-2s), & \frac{1}{2} + \frac{t}{2} \le s \le 1. \end{cases}$$

Again by 3.29, F is continuous. □

13.5 Lemma (reparametrization does not change homotopy class). *With a continuous map $\varphi : [0,1] \to [0,1]$, $\varphi(0) = 0$, $\varphi(1) = 1$, for any path α the path $\alpha \circ \varphi$ is path-homotopic to α.*

Proof. Let $F_t = (1-t)\varphi + t\mathrm{Id}_{[0,1]}$, a homotopy from φ to $\mathrm{Id}_{[0,1]}$ on $[0,1]$. We can check that $G_t = \alpha \circ F_t$, i.e., $G = \alpha \circ F$ is a path-homotopy from $\alpha \circ \varphi$ to α. □

13.6 Lemma. *$(\alpha \cdot \beta) \cdot \gamma$ is path-homotopic to $\alpha \cdot (\beta \cdot \gamma)$.*

Proof. We can check directly that $(\alpha \cdot \beta) \cdot \gamma$ is a reparametrization of $\alpha \cdot (\beta \cdot \gamma)$. Indeed,

$$\begin{aligned}(\alpha \cdot \beta) \cdot \gamma(t) &= \begin{cases} \alpha \cdot \beta(2t), & 0 \le t \le \frac{1}{2} \\ \gamma(2t-1), & \frac{1}{2} \le t \le 1 \end{cases} \\ &= \begin{cases} \alpha(4t), & 0 \le t \le \frac{1}{4} \\ \beta(4t-1), & \frac{1}{4} \le t \le \frac{1}{2} \\ \gamma(2t-1), & \frac{1}{2} \le t \le 1, \end{cases}\end{aligned}$$

and

$$\begin{aligned}\alpha \cdot (\beta \cdot \gamma)(t) &= \begin{cases} \alpha(2t), & 0 \le t \le \frac{1}{2} \\ \beta \cdot \gamma(2t-1) & \frac{1}{2} \le t \le 1 \end{cases} \\ &= \begin{cases} \alpha(2t), & 0 \le t \le \frac{1}{2} \\ \beta(4t-2), & \frac{1}{2} \le t \le \frac{3}{4} \\ \gamma(4t-3), & \frac{3}{4} \le t \le 1. \end{cases}\end{aligned}$$

We can see immediately that if $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma) \circ \varphi$ then φ has to be

$$\varphi(t) = \begin{cases} 2t, & 0 \le t \le \frac{1}{4} \\ t + \frac{1}{4}, & \frac{1}{4} \le t \le \frac{1}{2} \\ \frac{1}{2}t + \frac{1}{2}, & \frac{1}{2} \le t \le 1. \end{cases}$$

□

The fundamental group

Recall that a **group** is a set G with a map

$$\begin{aligned} G \times G &\to G \\ (a, b) &\mapsto a \cdot b, \end{aligned}$$

called a multiplication, such that:

(a) multiplication is associative: $(ab)c = a(bc)$, $\forall a, b, c \in G$,

(b) there exists a unique element of G called the identity element e satisfying $ae = ea = a$, $\forall a \in G$,

(c) for each element a of G there exists an element of G called the inverse element of a, denoted by a^{-1}, satisfying $a \cdot a^{-1} = a^{-1} \cdot a = e$.

See e.g. [Gal10] for more on groups.

Consider the set of loops of X based at a point x_0 under the path-homotopy relation. On this set we define a multiplication operation $[\alpha] \cdot [\beta] = [\alpha \cdot \beta]$. By 13.3 this operation is well-defined.

13.7 Theorem. *The set of all path-homotopy classes of loops of X based at a point x_0 is a group under the above operation.*

This group is called ***the* fundamental group** [1] of X at x_0, denoted by $\pi_1(X, x_0)$. The point x_0 is called the **base point**.

Proof. Denote by 1 the constant loop at x_0. Since $1 \cdot \alpha$ is a reparametrization of α via the map

$$\varphi(t) = \begin{cases} 0, & 0 \leq t \leq \frac{1}{2} \\ 2t - 1, & \frac{1}{2} \leq t \leq 1 \end{cases}$$

by lemma 13.5 $1 \cdot \alpha$ is path-homotopic to α, thus $[1] \cdot [\alpha] = [\alpha]$. So $[1]$ is the identity in $\pi_1(X, x_0)$.

Define $[\alpha]^{-1} = [\alpha^{-1}]$. It is easy to check that $[\alpha]^{-1}$ is well-defined, and is the inverse element of $[\alpha]$, by using 13.4.

By 13.6 we have associativity:

$$([\alpha] \cdot [\beta]) \cdot [\gamma] = [\alpha \cdot \beta] \cdot [\gamma] = [(\alpha \cdot \beta) \cdot \gamma] = [\alpha \cdot (\beta \cdot \gamma)] = [\alpha] \cdot ([\beta] \cdot [\gamma]).$$

□

Example. The fundamental group of a space containing only one point is trivial. Namely, if $X = \{x_0\}$ then $\pi_1(X, x_0) = 1$, the group with only one element.

Recall that a map $h : G_1 \to G_2$ between two groups is called a **group homomorphism** if it preserves group operations, i.e. $\forall a, b \in G_1, h(a \cdot b) = h(a) \cdot h(b)$. It is a **group isomorphism** if is a bijective homomorphism (in this case the inverse map is also a homomorphism).

13.8 Proposition (dependence on base point). *If there is a path from x_0 to x_1 then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, x_1)$.*

Proof. Let α be a path from x_0 to x_1. Consider the map

$$\begin{aligned} h_\alpha : \quad \pi_1(X, x_1) &\to \pi_1(X, x_0) \\ [\gamma] &\mapsto [\alpha \cdot \gamma \cdot \alpha^{-1}]. \end{aligned}$$

Using 13.3 and 13.6 we can check that this is a well-defined map, and is a group homomorphism with an inverse homomorphism:

[1] nhóm cơ bản

$$\begin{array}{rccl} h_\alpha^{-1}: & \pi_1(X,x_1) & \to & \pi_1(X,x_0) \\ & [\gamma] & \mapsto & [\alpha^{-1}\cdot\gamma\cdot\alpha]. \end{array}$$

□

Thus for a path-connected space the fundamental group is the same up to group isomorphisms for any choice of base point. Therefore if X is a path-connected space we often drop the base point in the notation and just write $\pi_1(X)$.

Induced homomorphism

Let X and Y be topological spaces, and $f : X \to Y$. Then f induces the following map

$$\begin{array}{rcl} f_*: \pi_1(X,x_0) & \to & \pi_1(Y,f(x_0)) \\ [\gamma] & \mapsto & [f\circ\gamma]. \end{array}$$

This is a well-defined map (problem 13.11). Notice that f_* depends on the base point x_0. Furthermore f_* is a group homomorphism, indeed:

$$f_*([\gamma_1]\cdot[\gamma_2]) = f_*([\gamma_1\cdot\gamma_2]) = [f\circ(\gamma_1\cdot\gamma_2)] = [f\circ\gamma_1]\cdot[f\circ\gamma_2] = f_*([\gamma_1])\cdot f_*([\gamma_2]).$$

Example. For $\mathrm{id} : X \to X$, the induced map is $\mathrm{id}_* = \mathrm{id}$. For a constant map $f : X \to \{x_0\}$, the induced map is $f_* = 1$.

Proposition. *If $f : X \to Y$ and $g : Y \to Z$ then $(g\circ f)_* = g_*\circ f_*$.*

Proof.

$$(g\circ f)_*([\gamma]) = [(g\circ f)\circ\gamma] = [g\circ(f\circ\gamma)] = g_*([f\circ\gamma)]) = g_*(f_*([\gamma])).$$

□

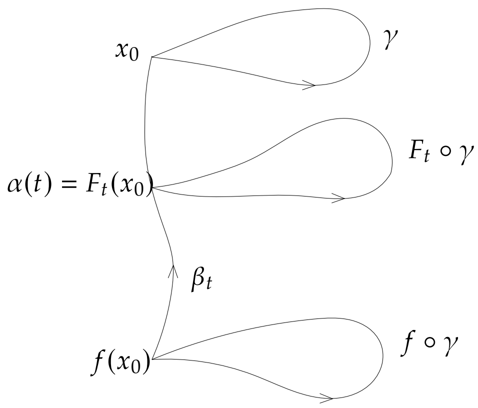
We deduce immediately that if $f : X \to Y$ is a homeomorphism then $f_* : \pi_1(X,x_0) \to \pi_1(Y,f(x_0))$ is a an isomorphism with inverse $(f_*)^{-1} = (f^{-1})_*$. This implies that ***the fundamental group is a topological invariant of path-connected spaces***.

Homotopy invariance

Lemma. *If $f : X \to X$ is homotopic to the identity map then $f_* : \pi_1(X,x_0) \to \pi_1(X,f(x_0))$ is an isomorphism.*

Proof. From the assumption there is a homotopy F from f to id_X. Then $\alpha(t) = F_t(x_0)$, $0\le t\le 1$, is a continuous path from $f(x_0)$ to x_0. We will show that

$f_* = h_\alpha$ where h_α is the map used in the proof of 13.8, which was shown there to be an isomorphism. Since for any loop γ at x_0 we have $f_*([\gamma]) = [f \circ \gamma]$ while $h_\alpha(\gamma) = [\alpha \cdot \gamma \cdot \alpha^{-1}]$, this is reduced to showing that $f \circ \gamma$ is path-homotopic to $\alpha \cdot \gamma \cdot \alpha^{-1}$.



For each fixed $0 \le t \le 1$, let β_t be a path which goes along α from $\alpha(0) = f(x_0)$ to $\alpha(t)$, for example given by $\beta_t(s) = \alpha(st)$, $0 \le s \le 1$. For any loop γ at x_0, let $G_t = \beta_t \cdot (F_t \circ \gamma) \cdot \beta_t^{-1}$. That G is continuous can be checked by writing down the formula for G explicitly. Then G gives a path-homotopy from $f \circ \gamma$ to $\alpha \cdot \gamma \cdot \alpha^{-1}$. □

Theorem. *If $f : X \to Y$ is a homotopy equivalence then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism.*

Proof. Since f is a homotopy equivalence there is $g : Y \to X$ such that $g \circ f$ is homotopic to id_X and $f \circ g$ is homotopic to id_Y. By the above lemma, the composition

$$\pi_1(X, x_0) \xrightarrow{f_*} \pi_1(Y, f(x_0)) \xrightarrow{g_*} \pi_1(X, g(f(x_0)))$$

is an isomorphism, which implies that g_* is surjective. Similarly the composition

$$\pi_1(Y, f(x_0)) \xrightarrow{g_*} \pi_1(X, g(f(x_0))) \xrightarrow{f'_*} \pi_1(Y, f(g(f(x_0))))$$

is an isomorphism, where f'_* denotes the induced homomorphism by f with base point $g(f(x_0))$. This implies that g_* is injective. Since g_* is bijective, from the first composition we see that f_* is bijective. □

Corollary (homotopy invariance). *If two path-connected spaces are homotopic then their fundamental groups are isomorphic.*

We say that for path-connected spaces, ***the fundamental group is a homotopy invariant***.

Example. The fundamental group of a contractible space is trivial.

Problems

13.9. Suppose that $f : X \to Y$ is continuous. Let γ_1 and γ_2 be two paths in X which are path-homotopic. Show that $f \circ \gamma_1$ and $f \circ \gamma_2$ are path-homotopic.

13.10. Show that if X_0 is a path-connected component of X and $x_0 \in X_0$ then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X_0, x_0)$.

13.11. Let X and Y be topological spaces, $f : X \to Y$, $f(x_0) = y_0$. Show that the induced map

$$\begin{aligned} f_* : \pi_1(X, x_0) &\to \pi_1(Y, y_0) \\ [\gamma] &\mapsto [f \circ \gamma] \end{aligned}$$

is a well-defined.

13.12. Suppose that Y is a retract of X via a retraction $r : X \to Y$. Let $i : Y \hookrightarrow X$ be the inclusion map. With $y_0 \in Y$, show that $r_* : \pi_1(X, y_0) \to \pi_1(Y, y_0)$ is surjective and $i_* : \pi_1(Y, y_0) \to \pi_1(X, y_0)$ is injective.

13.13. Show that

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0).$$

13.14. A space is said to be **simply connected**[1] if it is path-connected and any loop is path-homotopic to a constant loop. Show that a space is simply-connected if and only if it is path-connected and its fundamental group is trivial.

13.15. Show that any contractible space is simply connected.

13.16. Show that a path-connected space is simply connected if and only if all paths with same initial points and same terminal points are path-homotopic, in other words, there is exactly one path-homotopy class from one point to another point.

13.17. * Show that in a space X the following statements are equivalent:

(a) $\pi_1(X, x_0) = 1, \forall x_0 \in X$.

(b) Every continuous map $S^1 \to X$ is homotopic to a constant map.

(c) Every continuous map $S^1 \to X$ has a continuous extension to a map $D^2 \to X$.

[1] đơn liên

14 The fundamental group of the circle

The main result of this section is:

Theorem. *The fundamental group of the circle is infinite cyclic, that is,*

$$\pi_1(S^1) \cong \mathbb{Z}.$$

The remainder of this section devotes to explanation and proof of this theorem.

Let γ_n be the loop $(\cos(n2\pi t), \sin(n2\pi t))$, $0 \le t \le 1$, the loop on the circle S^1 based at the point $(1, 0)$ that goes n times around the circle at uniform speed in the counter-clockwise direction if $n > 0$ and in the clockwise direction if $n < 0$. Consider the map

$$\begin{aligned} \Phi : \mathbb{Z} &\to \pi_1(S^1, (1, 0)) \\ n &\mapsto [\gamma_n]. \end{aligned}$$

This map associates each integer n with the path-homotopy class of γ_n. We will show that Φ is a group isomorphism, where $\mathbb{Z}$ has the usual additive structure. This implies that the fundamental group of the circle is generated by a loop that goes once around the circle in the counter-clockwise direction, and the homotopy class of a loop in the circle corresponds to an integer representing the "number of times" that the loop goes around the circle, with the counter-clockwise direction being the positive direction.

Φ is a group homomorphism

This means γ_{m+n} is path-homotopic to $\gamma_m \cdot \gamma_n$. If $m = -n$ then we can verify from the formulas that $\gamma_n(t) = \gamma_{-m}(t) = \gamma_m(1-t) = \gamma_m^{-1}(t)$. We have checked earlier 13.4 that $\gamma_m \cdot \gamma_m^{-1}$ is path-homotopic to the constant loop γ_0.

If $m \neq -n$ then $\gamma_m \cdot \gamma_n$ is a reparametrization of γ_{m+n} so they are path-homotopic by 13.5. Indeed, with

$$\gamma_{m+n}(t) = (\cos((m+n)2\pi t), \sin((m+n)2\pi t)),\ 0 \le t \le 1,$$

and

$$\gamma_m \cdot \gamma_n(t) = \begin{cases} (\cos(m2\pi 2t), \sin(m2\pi 2t)), & 0 \le t \le \frac{1}{2}, \\ (\cos(n2\pi(2t-1)), \sin(n2\pi(2t-1))), & \frac{1}{2} \le t \le 1, \end{cases}$$

we can find φ such that $\gamma_{m+n} \circ \varphi = \gamma_m \cdot \gamma_n$, namely,

$$\varphi(t) = \begin{cases} \frac{m}{m+n} 2t, & 0 \le t \le \frac{1}{2}, \\ \frac{n(2t-1)+m}{m+n}, & \frac{1}{2} \le t \le 1. \end{cases}$$

Covering spaces

Let $p : \mathbb{R} \to S^1$, $p(t) = (\cos(2\pi t), \sin(2\pi t))$, a map that wraps the line around the circle countably infinitely many times in the counter-clockwise direction, called the projection map. This is related to the usual parametrization of the circle by angle. The map p is called the **covering map** associated with of the **covering space** [1] $\mathbb{R}$ of S^1. For a path $\gamma : [0,1] \to S^1$, a path $\tilde{\gamma} : [0,1] \to \mathbb{R}$ such that $p \circ \tilde{\gamma} = \gamma$ is called a ***lift*** of γ. For example $\gamma_n(t) = p(nt)$, $0 \le t \le 1$, so γ_n has a lift $\widetilde{\gamma_n}$ with $\widetilde{\gamma_n}(t) = nt$.

$$\begin{tikzcd} & \mathbb{R} \arrow[d, "p"] \\ {[0,1]} \arrow[ur, "\tilde{\gamma}"] \arrow[r, "\gamma"'] & S^1 \end{tikzcd}$$

A reason for considering covering spaces is that paths in S^1 are projections of paths in $\mathbb{R}$, while the homotopy of paths in $\mathbb{R}$ can be simpler than that in S^1.

As a demonstration, we use this idea of covering space to show again that γ_{m+n} is path-homotopic to $\gamma_m \cdot \gamma_n$. Consider the case $m + n > 0$. Let

$$\begin{aligned} \widetilde{\gamma_{m+n}} : [0,1] &\to [0, m+n] \\ t &\mapsto (m+n)t, \end{aligned}$$

then this is a lift of γ_{m+n}. Let

$$\begin{aligned} \widetilde{\gamma_m \cdot \gamma_n} : [0,1] &\to [0, m+n] \\ t &\mapsto \begin{cases} m2t, & 0 \le t \le \frac{1}{2}, \\ n(2t-1)+m, & \frac{1}{2} \le t \le 1, \end{cases} \end{aligned}$$

then this is a lift of $\gamma_m \cdot \gamma_n$. The two paths $\widetilde{\gamma_{m+n}}$ and $\widetilde{\gamma_m \cdot \gamma_n}$ are paths in $\mathbb{R}$ with same endpoints, so they are path-homotopic, via a path-homotopy such as $F_s = (1-s)\widetilde{\gamma_{m+n}} + s\widetilde{\gamma_m \cdot \gamma_n}$, $0 \le s \le 1$ (see 13.2). Then γ_{m+n} is path-homotopic to $\gamma_m \cdot \gamma_n$ via the path-homotopy $p \circ F$.

[1] không gian phủ

Φ is surjective

This means every loop γ on the circle based at $(1,0)$ is path-homotopic to a loop γ_n. Our argument is based on the fact, proved below at 14.1, that there is a path $\tilde{\gamma}$ on $\mathbb{R}$ starting at 0 which is a lift of γ. Then $\tilde{\gamma}(1)$ is an integer n. On $\mathbb{R}$ the path $\tilde{\gamma}$ is path-homotopic to the path $\widetilde{\gamma_n}$, via a path-homotopy F, so γ is path-homotopic to γ_n via the path-homotopy $p \circ F$.

14.1 Lemma (existence of lift). *Every path in S^1 has a lift to $\mathbb{R}$. Furthermore if the initial point of the lift is specified then the lift is unique.*

Proof. Let us write $S^1 = U \cup V$ with $U = S^1 \backslash \{(0,-1)\}$ and $V = S^1 \backslash \{(0,1)\}$. Then $p^{-1}(U) = \bigcup_{n\in\mathbb{Z}}(n-\frac{1}{4}, n+\frac{3}{4})$, consisting of infinitely many disjoint open subsets of $\mathbb{R}$, each of which is homeomorphic to U via p, i.e. $p : (n-\frac{1}{4}, n+\frac{3}{4}) \to U$ is a homeomorphism, in particular the inverse map exists and is continuous. The same happens with respect to V.

Let $\gamma : [0,1] \to S^1$, $\gamma(0) = (1,0)$. We can divide $[0,1]$ into sub-intervals with endpoints $0 = t_0 < t_1 < \cdots < t_n = 1$ such that on each sub-interval $[t_{i-1}, t_i]$, $1 \le i \le n$, the path γ is either contained in U or in V. This is guaranteed by the existence of a Lebesgue number (see 6.1) with respect to the open cover $\gamma^{-1}(U) \cup \gamma^{-1}(V)$ of $[0,1]$.

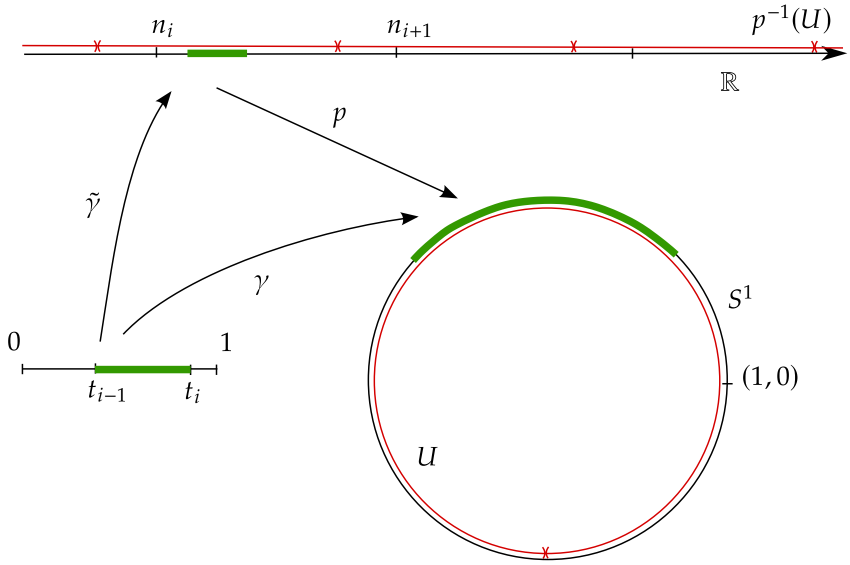


Figure 14.2: Construction of a lift of a path.

Suppose an initial point $\tilde{\gamma}(0)$ is chosen, an integer. Suppose that $\tilde{\gamma}$ has been uniquely extended to $[0, t_{i-1}]$ for a certain $1 \le i \le n$. If $\gamma([t_{i-1}, t_i]) \subset U$ then there is a unique $n_i \in \mathbb{Z}$ such that $\tilde{\gamma}(t_{i-1}) \in (n_i - \frac{1}{4}, n_i + \frac{3}{4})$. There is only one way to continuously extend $\tilde{\gamma}$ to $[t_{i-1}, t_i]$, that is by defining $\tilde{\gamma} = p|^{-1}_{(n_i-\frac{1}{4}, n_i+\frac{3}{4})} \circ \gamma$. See Fig. 14.2. In this way inductively $\tilde{\gamma}$ is extended continuously to $[0,1]$, uniquely. □

Examining the proof above we can see that the key property of the covering space $p : \mathbb{R} \to S^1$ is the following: each point on the circle has an open neighborhood U such that the preimage $p^{-1}(U)$ is the disjoint union of open subsets of $\mathbb{R}$, each of which is homeomorphic to U via p. This is the defining property of general covering spaces.

Φ is injective

This is reduced to showing that if γ_m is path-homotopic to γ_n then $m = n$. Our proof is based on another important result below, 14.3, which says that if γ_m is path-homotopic to γ_n then $\widetilde{\gamma_m}$ is path-homotopic to $\widetilde{\gamma_n}$. This implies the terminal point m of $\widetilde{\gamma_m}$ must be the same as the terminal point n of $\widetilde{\gamma_n}$.

14.3 Lemma (homotopy of lifts). *Lifts of path-homotopic paths with same initial points are path-homotopic.*

Proof. The proof is similar to the above proof of 14.1. Let $F : [0,1] \times [0,1] \to S^1$ be a path-homotopy from the path F_0 to the path F_1. If the two lifts $\widetilde{F_0}$ and $\widetilde{F_1}$ have same initial points then that initial point is the lift of the point $F((0,0))$.

As we noted earlier, the circle has an open cover O such that each $U \in O$ we have $p^{-1}(U)$ is the disjoint union of open subsets of $\mathbb{R}$, each of which is homeomorphic to U via p. The collection $\{F^{-1}(U) \mid U \in O\}$ is an open cover of the square $[0,1] \times [0,1]$. By the existence of Lebesgue number, there is a partition of $[0,1] \times [0,1]$ into sub-rectangles such that each sub-rectangle is contained in an element of $F^{-1}(O)$. More concisely, we can divide $[0,1]$ into sub-intervals with endpoints $0 = t_0 < t_1 < \cdots < t_n = 1$ such that for each $1 \leq i, j \leq n$ there is $U \in O$ such that $F([t_{i-1}, t_i] \times [t_{j-1}, t_j]) \subset U$.

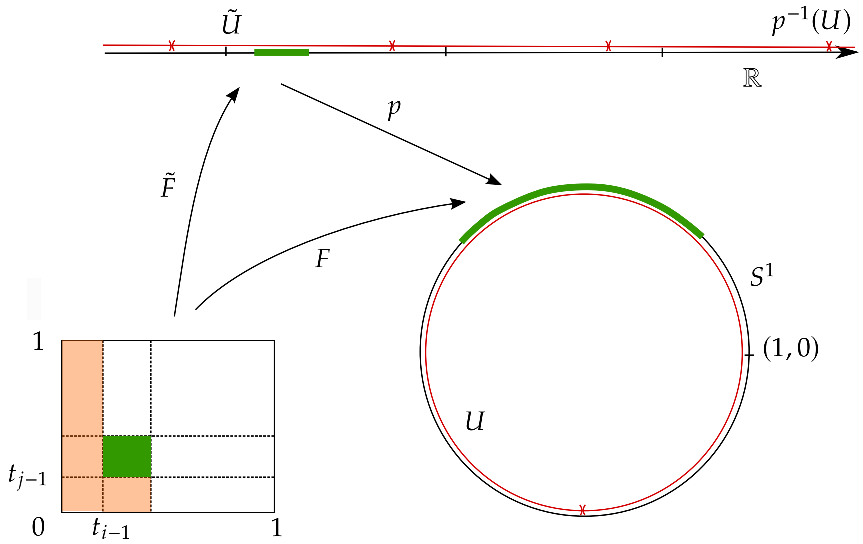


Figure 14.4: Construction of a lift of a path-homotopy.

We will build up $\tilde{F}$ sub-rectangle by sub-rectangle, going on columns from left to right and on each column going from bottom up. We already have $\tilde{F}((0,0))$. Suppose that now it is the turn of the sub-rectangle $[t_{i-1}, t_i]\times[t_{j-1}, t_j]$, for some $1 \le i, j \le n$, on which $\tilde{F}$ is to be extended to. We have $F([t_{i-1}, t_i] \times [t_{j-1}, t_j]) \subset U$ for a certain $U \in \mathcal{O}$. Let $\tilde{U}$ be the unique open subset of $\mathbb{R}$ such that $\tilde{U}$ contains the point $\tilde{F}((t_{i-1}, t_{j-1}))$ and $p|_{\tilde{U}} : \tilde{U} \to U$ is a homeomorphism. By the way we proceed one sub-rectangle at a time, the intersection A of the previous domain of $\tilde{F}$ and the sub-rectangle $[t_{i-1}, t_i] \times [t_{j-1}, t_j]$ is either one point, one edge, or the union of two edges with a common vertex, in all cases A is connected containing the vertex (t_{i-1}, t_{j-1}), therefore $\tilde{F}(A)$ is connected. Since $\tilde{F}(A) \subset p^{-1}(U)$, a disjoint union of open sets, and $\tilde{F}(A)\cap\tilde{U} \neq \emptyset$, we must have $\tilde{F}(A) \subset \tilde{U}$. We extend $\tilde{F}$ by defining $\tilde{F}$ on the sub-rectangle $[t_{i-1}, t_i] \times [t_{j-1}, t_j]$ to be $p|_{\tilde{U}}^{-1} \circ F$. This extension agrees with the previous function on the common domain A. See Fig. 14.4. The continuity of $\tilde{F}$ follows from gluing of continuous functions, see 3.29.

Thus we obtained a continuous lift $\tilde{F}$ of F. Since the initial point is given, by uniqueness of lifts of paths in 14.1, the restriction of $\tilde{F}$ to $[0,1]\times\{0\}$ is $\widetilde{F_0}$ while the restriction of $\tilde{F}$ to $[0,1]\times\{1\}$ is $\widetilde{F_1}$. Also by uniqueness of lifts, the lifts of F on $\{0\} \times [0,1]$ and $\{1\} \times [0,1]$ are constants. Thus $\tilde{F}$ is a path-homotopy from $\widetilde{F_0}$ to $\widetilde{F_1}$. □

Now we consider several applications of the theorem.

That the fundamental group of the circle is non-trivial gives us:

Corollary. *The circle is not contractible.*

14.5 Corollary. *There cannot be any retraction from the disk D^2 to its boundary S^1.*

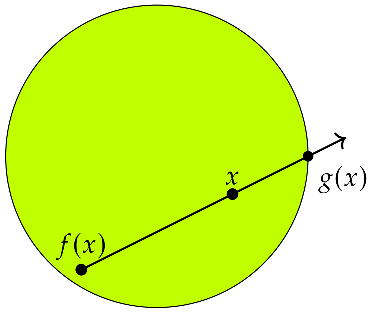
Proof. Suppose there is a retraction $r : D^2 \to S^1$. Let $i : S^1 \hookrightarrow D^2$ be the inclusion map. From the diagram $S^1 \xrightarrow{i} D^2 \xrightarrow{r} S^1$ we have $r \circ i = \mathrm{id}_{S^1}$, therefore on the fundamental groups $(r \circ i)_* = r_* \circ i_* = \mathrm{id}_{\pi_1(S^1)}$. A consequence is that r_* is onto, but this is not possible since $\pi_1(D^2)$ is trivial while $\pi_1(S^1)$ is non-trivial. □

A proof of this result for higher dimensions is presented in 17.6.

The important Brouwer fixed point theorem follows from that simple result:

14.6 Theorem (Brouwer fixed point theorem for dimension two). *Any continuous map from the disk D^2 to itself has a fixed point.*

Proof. Suppose that $f : D^2 \to D^2$ does not have a fixed point, i.e. $f(x) \neq x$ for all $x \in D^2$. The straight line from $f(x)$ to x will intersect the boundary ∂D^2 at a point $g(x)$.



We can check that g is continuous (see 14.10). Then $g : D^2 \to \partial D^2$ is a retraction, contradicting 14.5. □

Problems

14.7. Find the fundamental groups of:

(a) the Mobius band,

(b) the cylinder.

14.8. Show that the plane minus a point is not simply connected.

14.9. Compute the fundamental group of the torus.

14.10. Check that the map g in the proof of the Brouwer fixed point theorem is indeed continuous.

15 Van Kampen theorem

Van Kampen theorem is about giving the fundamental group of a union of subspaces from the fundamental groups of the subspaces.

Example ($S^1 \vee S^1$)**.** Two circles with one common point (the figure 8) is called a wedge product $S^1 \vee S^1$, see 8.19. In problem 15.7 we see that this is homotopic to the plane minus two points. Let x_0 be the common point, let a be a loop starting at x_0 going once around the first circle and let b the a loop starting at x_0 going once around the second circle. Then a and b generate the fundamental groups of the two circles with based points at x_0. Intuitively we can see that $\pi_1(S^1 \vee S^1, x_0)$ consists of path-homotopy classes of loops like a, ab, bba, $aabab^{-1}a^{-1}a^{-1}$, ... This is a group called the free group generated by a and b, denoted by $\langle a, b \rangle$.

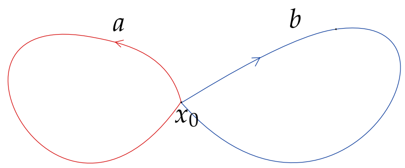


Figure 15.1: $S^1 \vee S^1$.

Free group

Let S be a set. Let S^{-1} be a disjoint set having a bijection with S. Corresponding to each element $x \in S$ is an element in S^{-1} denoted by x^{-1}. A **word with letters in S** is a finite sequence of elements in $S \cup S^{-1}$ (thus concisely it is a map from a finite subset of the set of positive integers to $S \cup S^{-1}$). The sequence with no element is called the **empty word**. Given two words we form a new word by juxtaposition (đặt kề): $(s_1 s_2 \cdots s_n) \cdot (s'_1 s'_2 \cdots s'_m) = s_1 s_2 \cdots s_n s'_1 s'_2 \cdots s'_m$. With this operation the set of all words with letters in S becomes a group. The identity element 1 is the empty word. The inverse element of a word $s_1 s_2 \cdots s_n$ is the word $s_n^{-1} s_{n-1}^{-1} \cdots s_1^{-1}$. This group is called the **free group generated by the set S**, denoted by $\langle S \rangle$.

Example. The free group $\langle \{a\} \rangle$ generated by the set $\{a\}$ is often written as $\langle a \rangle$. As a set $\langle a \rangle$ is often written as $\{a^n \mid n \in \mathbb{Z}\}$. The product is given by $a^m \cdot a^n = a^{m+n}$. The identity is a^0. Thus as a group $\langle a \rangle$ is an infinite cyclic group, isomorphic to $(\mathbb{Z}, +)$.

Let S be a set and let R be a set of some words with letters in S, i.e. R is a subset of the free group $\langle S \rangle$. Let N be the smallest normal subgroup of $\langle S \rangle$ containing R. The quotient group $\langle S \rangle / N$ is written $\langle S \mid R \rangle$. Elements of S are called **generators** of this group and elements of R are called ***relators*** of this group. We can think of $\langle S \mid R\rangle$ as consisting of words in S subjected to the **relations** $r = 1$ for all $r \in R$.

Example. $\left\langle a \mid a^2 \right\rangle = \{a^0, a^1\} \cong \mathbb{Z}_2$.

In a particular case, let R be the set of words of the form $aba^{-1}b^{-1}$ with $a \in S, b \in S$, then N is the free group generated by R, and $\langle S \mid R\rangle = \langle S\rangle /N = \langle S\rangle /\langle R\rangle$ is an abelian group, since the relation $aba^{-1}b^{-1} = 1$ means $ab = ba$. This group $\left\langle S \mid \forall a \in S, \forall b \in S, aba^{-1}b^{-1} \right\rangle$ is called the **free abelian group generated by** G.

Example. Consider the free abelian group generated by two elements a and b, $\left\langle a, b \mid aba^{-1}b^{-1} \right\rangle = \langle a, b \mid ab = ba\rangle$. Since a and b are commutative, each element of this group – a word in a, b, a^{-1}, b^{-1}– can be reduced to the form $a^m b^n$, where $m \in \mathbb{Z}, n \in \mathbb{Z}$. Thus $\langle a, b \mid ab = ba\rangle = \{a^m b^n \mid m \in \mathbb{Z}, n \in \mathbb{Z}\}$. For abelian group additive notation is often used, in that case each element of the free abelian group can be written as a integer linear combination of the generators. Thus each element of the free abelian group generated by a and b is of the form $ma + nb$ for some $m \in \mathbb{Z}, n \in \mathbb{Z}$.

Free product of groups

Let G and H be groups. Form the set of all words with letters in G or H. In such a word, two consecutive elements from the same group can be reduced by the group operation. For example $ba^2ab^3b^{-5}a^4 = ba^3b^{-2}a^4$. In particular if x and x^{-1} are next to each other then they will be canceled. The identity elements of G and H are also reduced. For example $abb^{-1}c = a1c = ac$.

As with free groups, given two words we form a new word by juxtaposition. For example $(a^2b^3a^{-1}) \cdot (a^3ba) = a^2b^3a^{-1}a^3ba = a^2b^3a^2ba$. This is a group operation, with the identity element 1 being the empty word, the inverse of a word $s_1s_2\cdots s_n$ is the word $s_n^{-1}s_{n-1}^{-1}\cdots s_1^{-1}$. This group is called the **free product** of G with H.

Example ($G * H \neq G \times H$). We have

$$\left\langle g \right\rangle * \langle h\rangle = \left\langle g, h \right\rangle = \{g^{m_1}h^{n_1}g^{m_2}h^{n_2}\cdots g^{m_k}h^{n_k} \mid m_1, n_1, \ldots, m_k, n_k \in \mathbb{Z},\ k \in \mathbb{Z}^+\}.$$

Compare that to $\left\langle g \right\rangle \times \langle h\rangle = \{(g^m, h^n) \mid m, n \in \mathbb{Z}\}$ with component-wise multiplication. This group can be identified with $\left\langle g, h \mid gh = hg \right\rangle = \{g^m h^n \mid m, n \in \mathbb{Z}\}$. Thus $\mathbb{Z} * \mathbb{Z} \neq \mathbb{Z} \times \mathbb{Z}$.

For more details on free group and free product, see textbooks on Algebra such as [Gal10] or [Hun74].

Van Kampen theorem

The following is our main tool for computing the fundamental groups:

Theorem **(Van Kampen theorem).** *Suppose that $U, V \subset X$ are open, path-connected, $U \cap V$ is path-connected, and $x_0 \in U \cap V$. Let $i_U : U \cap V \hookrightarrow U$ and $i_V : U \cap V \hookrightarrow V$ be inclusion maps. Then*

$$\pi_1(U \cup V, x_0) \cong \frac{\pi_1(U, x_0) * \pi_1(V, x_0)}{\langle \{(i_U)_*(\alpha)(i_V)_*(\alpha)^{-1} \mid \alpha \in \pi_1(U \cap V, x_0)\} \rangle}.$$

Very roughly, counting up to path-homotopies, the loops in $U \cup V$ consist of loops in U join with loops in V, however loops in $U \cap V$ should be counted only once.

The theorem can be formulated for coverings with more than two elements. It is also called the Seifert–Van Kampen theorem.

Proof. Let γ be any loop in $U \cup V$ at x_0. By the existence of a Lebesgue number in association with the open cover $\{\gamma^{-1}(U), \gamma^{-1}(V)\}$ of $[0,1]$, there is a partition of $[0,1]$ by $0 = t_0 < t_1 < \cdots < t_n = 1$ such that for every $1 \leq i \leq n$ either $\gamma([t_{i-1}, t_i]) \subset U$ or $\gamma([t_{i-1}, t_i]) \subset V$, furthermore we can arrange so that $\gamma(t_i) \in U \cap V$. Let γ_i be the path $\gamma\big|_{[t_{i-1}, t_i]}$ reparametrized to the domain $[0,1]$. Then γ has a reparametrization as $\gamma_1 \cdot \gamma_2 \cdots \gamma_n$. Let β_i be a path in $U \cap V$ from $\gamma(t_i)$ to x_0, $1 \leq i \leq n-1$. In terms of path-homotopy in $U \cup V$,

$$\begin{aligned} \gamma &\sim \gamma_1 \cdot \gamma_2 \cdots \gamma_n \\ &\sim (\gamma_1 \cdot \beta_1) \cdot (\beta_1^{-1} \cdot \gamma_2 \cdot \beta_2) \cdots (\beta_{n-2}^{-1} \cdot \gamma_{n-1} \cdot \beta_{n-1})(\beta_{n-1}^{-1} \cdot \gamma_n). \end{aligned}$$

Thus every loop at x_0 in $U \cup V$ is path-homotopic to a product of loops at x_0 each of which is contained entirely in either U or V.

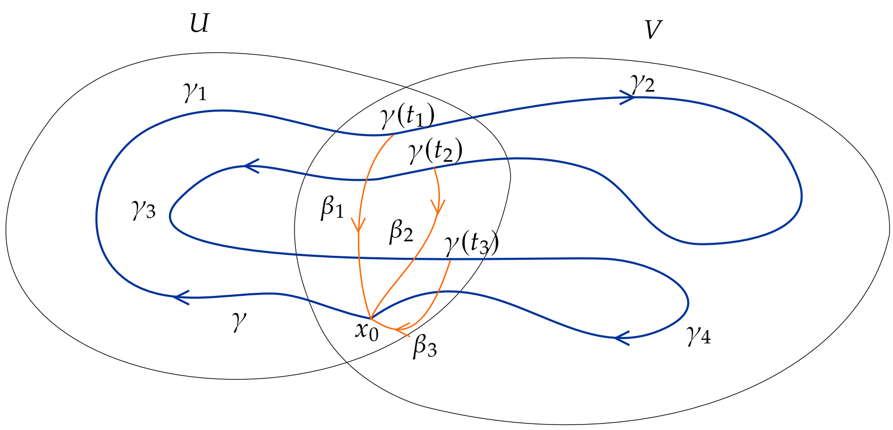


Figure 15.2: Ideas for a proof of Van Kampen theorem.

Let $j_U : U \hookrightarrow U \cup V$ and $j_V : V \hookrightarrow U \cup V$ be inclusion maps. Let

$$\begin{aligned} \Phi : \pi_1(U, x_0) * \pi_1(V, x_0) &\to \pi_1(U \cup V, x_0) \\ a_1 b_1 \cdots a_n b_n &\mapsto (j_U)_*(a_1)\,(j_V)_*(b_1) \cdots (j_U)_*(a_n)\,(j_V)_*(b_n). \end{aligned}$$

The above discussion implies that Φ is surjective. It is also immediate that Φ is a homomorphism.

For the kernel of Φ, let $\alpha \in \pi_1(U \cap V, x_0)$, then

$$\begin{aligned} \Phi((i_U)_*(\alpha)(i_V)_*(\alpha)^{-1}) &= (j_U)_*((i_U)_*(\alpha))\,(j_V)_*((i_V)_*(\alpha)^{-1})) \\ &= (j_U \circ i_U)_*(\alpha)\,(j_V \circ i_V)_*(\alpha)^{-1} = 1 \end{aligned}$$

since $j_U \circ i_U = j_V \circ i_V$. Thus $\ker \Phi$ contains the normal subgroup generated by all elements $(i_U)_*(\alpha)(i_V)_*(\alpha)^{-1}$, $\alpha \in \pi_1(U \cap V, x_0)$.

That $\ker \Phi$ is equal to that group is more difficult, we do not present a proof, instead refer to [Vic94]. □

Corollary. *If $X = U \cup V$ with U, V open, path-connected, $U \cap V$ is simply connected, and $x_0 \in U \cap V$, then $\pi_1(X, x_0) \cong \pi_1(U, x_0) * \pi_1(V, x_0)$.*

Corollary. *The spheres of dimensions greater than one are simply connected:*

$$\pi_1(S^n) \cong \begin{cases} \mathbb{Z}, & n = 1 \\ 1, & n > 1. \end{cases}$$

Proof. Let $A = S^n \setminus \{(0, 0, \ldots, 0, 1)\}$ and $B = S^n \setminus \{(0, 0, \ldots, 0, -1)\}$. Then A and B are contractible. If $n \geq 2$ then $A \cap B$ is path-connected. By Van Kampen theorem, $\pi_1(S^2) \cong \pi_1(A) * \pi_1(B) = 1$. □

Example ($S^1 \vee S^1$). Let U be the union of the first circle with an open arc on the second circle containing the common point. Similarly let V be the union of the second circle with an open arc on the first circle containing the common point. See Fig. 15.1. Clearly U and V have deformation retractions to the first and the second circles respectively, while $U \cap V$ is simply connected, having a deformation retraction to the common point. Applying the Van Kampen theorem,

$$\pi_1(S^1 \vee S^1) \cong \pi_1(S^1) * \pi_1(S^1) \cong \mathbb{Z} * \mathbb{Z}.$$

The fundamental group of a cell complex

15.3 Theorem. *Let X be a path-connected topological space and consider the space $X \sqcup_f D^n$ obtained by attaching an n-dimensional cell to X via the map $f : \partial D^n = S^{n-1} \to X$.*

(a) If $n > 2$ then $\pi_1(X \sqcup_f D^n) \cong \pi_1(X)$.

(b) If $n = 2$, let γ_1 be the loop $(\cos 2\pi t, \sin 2\pi t)$ on the circle ∂D^2, let $x_0 = f(1, 0)$, then $\pi_1(X \sqcup_f D^n, x_0) \cong \pi_1(X, x_0)/\langle[f \circ \gamma_1]\rangle$.

Intuitively, ***gluing a disk of dimension 2 destroys the homotopy class of the boundary circle of the disk***, while ***gluing a disk of dimension greater than 2 does not affect the fundamental group***.

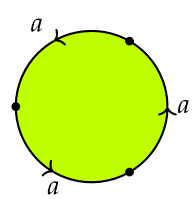
Proof. Let $Y = X \sqcup_f D^n$. Let $U = X \sqcup_f \{x \in D^n \mid \|x\| > \frac{1}{2}\} \subset Y$. There is a deformation retraction from U to X. Let V be the image of the embedding of the interior of D^n in Y. Then V is contractible. Also, $U \cap V$ is homeomorphic to $\{x \in D^n \mid \frac{1}{2} < \|x\| < 1\}$, which has a deformation retraction to S^{n-1}. We now apply Van Kampen theorem to the pair (U, V).

When $n > 2$ the fundamental group of $U \cap V$ is trivial, therefore $\pi_1(Y) \cong \pi_1(U) \cong \pi_1(X)$.

Consider the case $n = 2$. Let $y_0 \in U \cap V$ be the image of the point $(\frac{2}{3}, 0) \in D^2$, let γ be imbedding of the loop $\frac{2}{3}(\cos 2\pi t, \sin 2\pi t)$ starting at y_0 going once around the annulus $U \cap V$. Then $[\gamma]$ is a generator of $\pi_1(U \cap V, y_0)$. In V the loop γ is homotopically trivial, therefore $\pi_1(Y, y_0) \cong \pi_1(U, y_0)/\langle[\gamma]\rangle$. Under the deformation retraction from U to X, $[\gamma]$ becomes $[f \circ \gamma_1]$. Therefore $\pi_1(Y, x_0) \cong \pi_1(X, x_0)/\langle[f \circ \gamma_1]\rangle$. □

This result shows that the fundamental group only gives information about the two-dimensional skeleton of a cell complex, it does not give information on cells of dimensions greater than 2.

15.4 Example. Consider the space below.



As a cell-complex it consists of one 0-cell, one 1-cell (represented by a), forming the 1-dimensional skeleton (a circle, also denoted by a), and one 2-cell attached to the 1-dimensional skeleton by wrapping the boundary of the disk around a three times. Thus the fundamental group of the space is isomorphic to $\langle a \mid a^3 = 1\rangle \cong \mathbb{Z}_3$. In practice, we get the fundamental group immediately by arguing that the boundary of the disk is a^3 and gluing this disk homotopically destroys its boundary.

Fundamental groups of surfaces

By the classification theorem, any compact without boundary surface is obtained by identifying the edges of a polygon following a word as in 11.1. As such it has a cell complex structure with a two-dimensional disk glued to the boundary of the polygon under the equivalence relation, which is a wedge of circles. An application of 15.3 gives us:

15.5 Theorem. *The fundamental group of a connected compact surface S is isomorphic to one of the following groups:*

(a) *trivial group, if $S = S^2$,*

(b) $\left\langle a_1, b_1, a_2, b_2, \dots, a_g, b_g \mid a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1} \right\rangle$*, if S is the orientable surface of genus g,*

(c) $\left\langle c_1, c_2, \dots, c_g \mid c_1^2 c_2^2 \cdots c_g^2 \right\rangle$*, if S is the un-orientable surface of genus g.*

This is Problem 15.11.

Example. For the torus, $\pi_1(T_1) \cong \langle a, b \mid ab = ba \rangle \cong \mathbb{Z} \times \mathbb{Z}$.
For the projective plane, $\pi_1(\mathbb{R}P^2) \cong \left\langle c \mid c^2 = 1 \right\rangle \cong \mathbb{Z}_2$.

For more examples and topics on the fundamental group the reader may see [Hat01]. From Theorem 15.3 it is not surprising that the fundamental group features prominently in topology of low dimensional manifolds, see for instance [Dale Rolfsen, *Knots and Links*, Publish or Perish, 1990].

Problems

15.6. Use the Van Kampen theorem to find the fundamental groups of the following spaces:

(a) A wedge of finitely many circles.

(b) $S^1 \vee S^2$.

(c) $S^2 \vee S^3$.

15.7. Show that the plane minus finitely many points has a deformation retraction to a bouquet of circles, then find its fundamental group.

15.8. Find the fundamental group of the Euclidean space $\mathbb{R}^3$ minus finitely many points.

15.9. By Problem 8.34 we have two different presentations for the fundamental group of the Klein bottle: $\pi_1(K) \cong \left\langle a, b \mid aba^{-1}b \right\rangle$ and $\pi_1(K) \cong \left\langle a, b \mid a^2 b^2 \right\rangle$. Show directly that $\left\langle a, b \mid aba^{-1}b \right\rangle \cong \left\langle a, b \mid a^2 b^2 \right\rangle$.

15.10. Is the fundamental group of the Klein bottle abelian?

15.11. Give details for a proof of Theorem 15.5.

15.12. Show that the fundamental groups of the one-hole torus and the two-holes torus are not isomorphic. Deduce that the two surfaces are not homeomorphic.

15.13. Find a space whose fundamental group is isomorphic to $\mathbb{Z} * \mathbb{Z}_2 * \mathbb{Z}_3$.

15.14. Show that if $X = A \cup B$, where A and B are open, simply connected, and $A \cap B$ is path-connected, then X is simply connected.

15.15. Is $\mathbb{R}^3 \setminus \{(0,0,0)\}$ simply connected? How about $\mathbb{R}^3 \setminus \{(0,0,0),(1,0,0)\}$?

15.16. Show that if a compact surface is simply connected then it is homeomorphic to the two-dimensional sphere.

15.17. Consider the three-dimensional torus T^3, obtained from the cube $[0,1]^3$ by identifying opposite faces by projection maps, that is $(x,y,0) \sim (x,y,1)$, $(0,y,z) \sim (1,y,z)$, $(x,0,z) \sim (x,1,z)$, $\forall (x,y,z) \in [0,1]^3$.

(a) Show that T^3 is homeomorphic to $S^1 \times S^1 \times S^1$.

(b) Show that T^3 is a 3-dimensional manifold.

(c) Construct a cellular structure on T^3.

(d) Compute the fundamental group of T^3 from the cell structure.

15.18. * Consider the space X obtained from the cube $[0,1]^3$ by rotating each lower face (i.e. faces on the planes xOy, yOz, zOx) an angle of $\pi/2$ about the normal line that goes through the center of the face, then gluing this face to the opposite face. For example the point (1,1,0) is glued to the point $(0,1,1)$.

(a) Show that X has a cellular structure consisting of 2 0-cells, 4 1-cells, 3 2-cells, 1 3-cell.

(b) Compute the fundamental group of X.

(c) Check that the fundamental group of X is isomorphic to the quaternion group.

15.19. If

$$G = \left\langle g_1, g_2, \ldots, g_{m_1} \mid r_1, r_2, \ldots, r_{n_1} \right\rangle$$

and

$$H = \langle h_1, h_2, \ldots, h_{m_2} \mid s_1, s_2, \ldots, s_{n_2} \rangle$$

find $G * H$.

16 Simplicial homology

In homology theory we associate each space (with certain additional structures) with a sequence of groups.

Oriented simplex

On a simplex of dimension greater than 0, consider the relation on the collection of ordered sets of vertices of this simplex whereas two order sets of vertices are related if they differ by an even permutation. This is an equivalence relation. Each of the two equivalence classes is called an **orientation** of the simplex. If we choose an orientation, then the simplex is said to be **oriented**.

In more details, an orientation for a set A consisting of n elements is represented by bijective map from the set of integers $I = \{1, 2, \ldots, n\}$ to A. If o and o' are two such maps, then $o = o' \circ \sigma$ where $\sigma = o'^{-1} \circ o : I \to I$ is bijective, a permutation. Any permutation falls into one of two classes: even or odd permutations. We say that o and o' represent the same orientations if σ is an even permutation.

For each simplex of dimension greater than 0 there are two oriented simplices. For convenience we say that for a 0-dimensional simplex (a vertex) there is only one orientation.

Remark. In this section, to be clearer we reserve the notation $v_0 v_1 \cdots v_n$ for an un-oriented simplex with the set of vertices $\{v_0, v_1, \ldots, v_n\}$, and the notation $[v_0, v_1, \ldots, v_n]$ for an oriented simplex with the same set of vertices, i.e. the simplex $v_0 v_1 \cdots v_n$ with this particular order of vertices. (Later, and in other sources, this convention can be dropped.)

Example. A 1-dimensional simplex in $\mathbb{R}^n$ is a straight segment connecting two points. Choosing one point as the first point and the other point as the second gives an orientation to this simplex. Intuitively, this means to give a direction to the straight segment. If the two vertices are labeled v_0 and v_1, then an ordered pair $[v_0, v_1]$ gives an orientation for the simplex $v_0 v_1$, while an ordered pair $[v_1, v_0]$ gives an oriented simplex with the opposite orientation.

Consider a 2-dimensional simplex. Let v_0, v_1, v_2 be the vertices. The oriented simplices $[v_0, v_1, v_2]$, $[v_1, v_2, v_0]$, $[v_2, v_0, v_1]$ have same orientations, opposite to the orientations of the oriented simplices $[v_1, v_0, v_2]$, $[v_2, v_1, v_0]$, $[v_0, v_2, v_1]$.

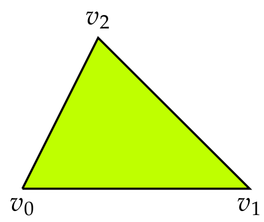


Figure 16.1

Chain

Let X be a simplicial complex in a Euclidean space. For each integer n let $S_n(X)$ be the free abelian group generated by all n-dimensional oriented simplices of X modulo the relation that two orientated simplices with opposite orientations are opposite in the group. Thus for each $\sigma \in X$ and $\dim \sigma = n$ there are two corresponding oriented simplices σ' and σ'' with opposite orientations, both σ' and σ'' are generators of $S_n(X)$, with a relation $\sigma' = -\sigma''$ (using additive notation).

Each element of $S_n(X)$ is an equivalence class, called an n-dimensional **chain** [1], is represented by a finite sum of integer multiples of n-dimensional oriented simplices, an integer linear combinations of oriented simplices, i.e. of the form $\sum_{i=1}^{m} n_i\sigma_i$ where σ_i is an n-dimensional oriented simplex of X and $n_i \in \mathbb{Z}$. We often drop the notation for equivalence class and use a representative to indicate its equivalence class.

If $n < 0$ or $n > \dim X$ then $S_n(X)$ is assigned to be the trivial group 0.

Example. Consider the simplicial complex whose underlying space is a line segment. So $X = \{AB, A, B\}$. The group $S_0(X)$ is generated by A and B. The following is a 0-chain: $2A - 3B$.

The group $S_1(X)$ is generated by $[A, B]$ and $[B, A]$ together with the relation $[B, A] = -[A, B]$. An example of a 1-chain is $2[A, B] + 5[B, A]$, but since $[B, A] = -[A, B]$, we have $2[A, B] + 5[B, A] = 2[A, B] - 5[A, B] = -3[A, B]$. We see that we can identify $S_1(X)$ with the group generated by $[A, B]$ and write an element of $S_1(X)$ as $n[A, B]$ with $n \in \mathbb{Z}$.

16.2 Example. Consider the simplicial complex whose underlying space is the triangle in figure 16.1:

$$X = \{v_0v_1v_2, v_0v_1, v_1v_2, v_2v_0, v_0, v_1, v_2\}.$$

For $n = 0$:

$$S_0(X) = \{n_0v_0 + n_1v_1 + n_2v_2 \mid n_0, n_1, n_2 \in \mathbb{Z}\}.$$

[1] xích, dây chuyền

For $n = 1$: In the group $S_1(X)$ we have the relations $[v_1, v_0] = -[v_0, v_1]$, $[v_2, v_1] = -[v_1, v_2]$, $[v_0, v_2] = -[v_2, v_0]$. We write

$$S_1(X) = \{n_0[v_0, v_1] + n_1[v_1, v_2] + n_2[v_2, v_0] \mid n_0, n_1, n_2 \in \mathbb{Z}\}.$$

For $n = 2$: Here for example $[v_0, v_1, v_2] = -[v_2, v_1, v_0] = [v_1, v_2, v_0]$. We can write

$$S_2(X) = \{n_0[v_0, v_1, v_2] \mid n_0 \in \mathbb{Z}\}.$$

To help understanding here is an alternative way to define a chain. An n-dimensional chain on a simplicial complex X is a map c from the set of all oriented n-dimensional simplices of X to the set of all integer numbers satisfying that if σ' and σ'' are oriented simplices with opposite orientations then $c(\sigma') = -c(\sigma'')$ and $c(\sigma') = 0$ for all but finitely many σ'. Then $S_n(X)$ is the free abelian group generated by all n-dimensional chains on X.

Although the group $S_n(X)$ is isomorphic to the free abelian group generated by the set of all n-dimensional simplices (without orientation) of X, we need orientation in order to discuss the boundary operator.

Boundary

Let $\sigma = [v_0, v_1, \ldots, v_n]$ be an n-dimensional oriented simplex. Define the **boundary** of σ to be the following $(n-1)$-dimensional chain, the alternating sum of the $(n-1)$-dimensional faces of σ:

$$\partial_n \sigma = \sum_{i=0}^{n} (-1)^i [v_0, v_1, \ldots, v_{i-1}, \widehat{v_i}, v_{i+1}, \ldots, v_n],$$

where the notation $\widehat{v_i}$ is traditionally used to indicate that this element is dropped. This map is extended linearly to become a map from $S_n(X)$ to $S_{n-1}(X)$, namely

$$\partial_n \left(\sum_{i=1}^{m} n_i \sigma_i \right) = \sum_{i=1}^{m} n_i \partial_n (\sigma_i).$$

We can check easily that $\partial_n(-\sigma) = -\partial_n(\sigma)$, thus the map ∂_n, which is defined on oriented simplices, induces a well-defined map on $S_n(X)$, and is a group homomorphism from $S_n(X)$ to $S_{n-1}(X)$.

Remark. If $n < 0$ or $n > \dim X$ then $\partial_n = 0$ since $S_n(X) = 0$.

Example. Continuing Example 16.2:

$$\partial_1([v_0, v_1]) = v_1 - v_0,$$

$$\partial_1([v_1, v_2]) = v_2 - v_1,$$

$$\partial_1([v_2, v_0]) = v_0 - v_2,$$

$$\partial_2([v_0, v_1, v_2]) = [v_1, v_2] - [v_0, v_2] + [v_0, v_1] = [v_1, v_2] + [v_2, v_0] + [v_0, v_1].$$

Notice that:

$$\begin{aligned} \partial_1(\partial_2([v_0, v_1, v_2])) &= \partial_1([v_0, v_1] + [v_1, v_2] + [v_2, v_0]) \\ &= (v_1 - v_0) + (v_2 - v_1) + (v_0 - v_2) = 0. \end{aligned}$$

16.3 Proposition (boundary of boundary is zero). $\partial_{n-1} \circ \partial_n = 0$.

This says intuitively "boundary has empty boundary":

Proof. Let $\sigma = [v_0, v_1, \ldots, v_n]$, an oriented n-simplex. By definition,

$$\partial_n(\sigma) = \sum_{i=0}^{n} (-1)^i [v_0, v_1, \ldots, v_{i-1}, \widehat{v_i}, v_{i+1}, \ldots, v_n].$$

Then

$$\begin{aligned} \partial_{n-1}\partial_n(\sigma) &= \sum_{i=0}^{n} (-1)^i \partial_{n-1}([v_0, v_1, \ldots, v_{i-1}, \widehat{v_i}, v_{i+1}, \ldots, v_n]) \\ &= \sum_{i=0}^{n} (-1)^i (\sum_{j=0}^{i-1} (-1)^j [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_n] + \\ &\quad + \sum_{j=i}^{n-1} (-1)^j [v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_{j+1}}, \ldots, v_n]) \\ &= \sum_{0 \le j < i \le n} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_n] + \\ &\quad + \sum_{0 \le i \le j \le n-1} (-1)^{i+j} [v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_{j+1}}, \ldots, v_n] \\ &= \sum_{0 \le j < i \le n} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_n] + \\ &\quad + \sum_{0 \le i < k \le n,\ (k=j+1)} (-1)^{i+k-1} [v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_k}, \ldots, v_n] \\ &= \sum_{0 \le j < i \le n} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_n] + \\ &\quad + \sum_{0 \le j < i \le n} -(-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_n] \\ &= 0. \end{aligned}$$

□

That $\partial_{n-1} \circ \partial_n = 0$ can be interpreted as

$$\operatorname{Im}(\partial_{n+1}) \subset \ker(\partial_n).$$

Elements of $\ker(\partial_n)$ are often called **closed chains** or **cycles** [1], while elements of $\mathrm{Im}(\partial_{n+1})$ are called **exact chains** or **boundaries**. We have just observed that

exact $\Rightarrow$ closed

boundaries are cycles

Homology

In general, a sequence of groups and homomorphisms

$$\cdots \xrightarrow{\partial_{n+2}} S_{n+1} \xrightarrow{\partial_{n+1}} S_n \xrightarrow{\partial_n} S_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} S_0$$

satisfying $\mathrm{Im}(\partial_{n+1}) \subset \ker(\partial_n)$, i.e. $\partial_n \circ \partial_{n+1} = 0$ for all $n \geq 0$ is called a **chain complex** [2]. If furthermore $\mathrm{Im}(\partial_{n+1}) = \ker(\partial_n)$, $\forall n \geq 0$ then the chain complex is called **exact** [3].

Notice that if the groups S_n are abelian then $\mathrm{Im}(\partial_{n+1})$ is a normal subgroup of $\ker(\partial_n)$.

The n-dimensional **simplicial homology group** [4] of a simplicial complex X is defined to be the quotient group

$$H_n(X) = \frac{\ker(\partial_n)}{\mathrm{Im}(\partial_{n+1})}.$$

Briefly:

$$H_n = \text{closed } n\text{-chains}/\text{exact } n\text{-chains} = n\text{-cycles}/n\text{-boundaries}.$$

Thus ***homology measures the set of cycles which are not boundaries***. Even more intuitively, a cycle which is not a boundary is a "hole" in space, so ***n-dimensional homology counts n-dimensional holes***.

Example. If $n < 0$ or $n > \dim X$ then $H_n(X) = 0$ since $S_n(X) = 0$. If $n = 0$ then $H_0(X) = S_0(X)/\mathrm{Im}\,\partial_1$. If $n = \dim X$ then $H_n(X) = \ker \partial_n$.

Example. Continuing example 16.2, we compute the homology groups of X.

From

$$\partial_1(n_0[v_0, v_1] + n_1[v_1, v_2] + n_2[v_2, v_0]) \quad = \quad n_0(v_1 - v_0) + n_1(v_2 - v_1) + n_2(v_0 - v_2),$$

[1] chu trình
[2] phức xích
[3] khớp
[4] nhóm đồng điều

we see that an element of $\operatorname{Im}\partial_1$ has the form

$$\begin{aligned} n_0(v_1-v_0)+n_1(v_2-v_1)+n_2(v_0-v_2) &= n_0(v_1-v_0)+n_1(v_2-v_0+v_0-v_1)+n_2(v_0-v_2) \\ &= (n_0-n_1)(v_1-v_0)+(n_1-n_2)(v_2-v_0). \end{aligned}$$

Thus $\operatorname{Im}\partial_1 = \langle\{v_1-v_0, v_2-v_0\}\rangle$. So

$$\begin{aligned} H_0(X) &= S_0(X)/\operatorname{Im}\partial_1 \cong \langle v_0, v_1, v_2 \mid v_1 = v_0, v_2 = v_0\rangle \\ &\cong \langle v_0\rangle \cong \langle v_1\rangle \cong \langle v_2\rangle \cong \mathbb{Z}. \end{aligned}$$

Since

$$n_0(v_1-v_0)+n_1(v_2-v_1)+n_2(v_0-v_2) = (n_2-n_0)v_0+(n_0-n_1)v_1+(n_1-n_2)v_2$$

we see that $\partial_1(n_0[v_0,v_1]+n_1[v_1,v_2]+n_2[v_2,v_0]) = 0$ if and only if $n_0 = n_1 = n_2$. So

$$\ker\partial_1 = \{n([v_0,v_1]+[v_1,v_2]+[v_2,v_0]) \mid n \in \mathbb{Z}\} = \langle([v_0,v_1]+[v_1,v_2]+[v_2,v_0])\rangle\,.$$

Since

$$\partial_2(n[v_0,v_1,v_2]) = n([v_0,v_1]+[v_1,v_2]+[v_2,v_0])$$

we see that

$$\operatorname{Im}\partial_2 = \{n([v_0,v_1]+[v_1,v_2]+[v_2,v_0]) \mid n \in \mathbb{Z}\} = \langle([v_0,v_1]+[v_1,v_2]+[v_2,v_0])\rangle\,.$$

Thus $\operatorname{Im}\partial_2 = \ker\partial_1$, so

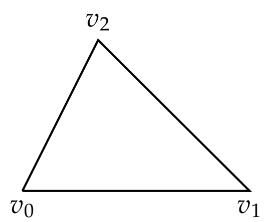
$$H_1(X) = 0.$$

We also see immediately that $\partial_2(n[v_0,v_1,v_2]) = 0$ if and only if $n = 0$. Thus $\ker\partial_2 = 0$ and

$$H_2(X) = \ker\partial_2 = 0.$$

Example. Instead of a 2-dimensional triangle in example 16.2, let us consider a 1-dimensional triangle, namely

$$X = \{v_0v_1, v_1v_2, v_2v_0, v_0, v_1, v_2\}.$$



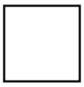
In this case $H_1(X) = \ker\partial_1 = \langle([v_0,v_1]+[v_1,v_2]+[v_2,v_0])\rangle \cong \mathbb{Z}$, $H_0(X) = \langle v_0\rangle \cong \mathbb{Z}$, and $H_n(X) = 0$ for $n \neq 0, 1$.

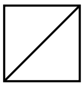
For more on simplicial homology one can read [Mun84], [Cro78], or [Ams83]. In [Hat01] Hatcher used a modified notion called Δ-complex, different from simplicial complex. For algorithms for computation of simplicial homology, see [KMM04, chapter 3].

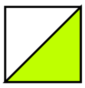
Problems

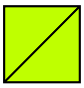
16.4. Compute the homology of the simplicial complex described in the figure:

(a)

(b) 

(c) 

(d) 

(e) 

16.5. Compute the simplicial homology group $H_2(X)$ where X is the simplicial complex representing a tetrahedron in $\mathbb{R}^3$, namely

$$X = \{v_0, v_1, v_2, v_3, v_0v_1, v_1v_2, v_2v_0, v_3v_0, v_3v_1, v_3v_2, v_0v_1v_2, v_1v_2v_3, v_2v_3v_0, v_0v_1v_3\}.$$

16.6. Show that if the underlying space of a simplicial complex is connected then for any two vertices there is a continuous path on the edges connecting the two vertices.

16.7. Given a simplicial complex X.

(a) Show that if the underlying space $|X|$ is connected then $H_0(X) \cong \langle v \rangle \cong \mathbb{Z}$ where v is any vertex of X.

(b) Show that if $|X|$ has k connected components then $H_0(X) \cong \bigoplus_{i=1}^{k} \langle v_i \rangle \cong \mathbb{Z}^k$, where v_i, $1 \le i \le k$ are arbitrary vertices in different components of $|X|$.

16.8. Show that if the underlying space of a simplicial complex X has k connected components X_i, $1 \le i \le k$, then for any $n \in \mathbb{Z}$ we have $H_n(X) \cong \bigoplus_{i=1}^{k} H_n(X_i)$.

16.9. As we may observe, thanks to finite structures it is possible to compute simplicial homology using computer programs. One such program is `GAP` [1] with packages `simpcomp` or `HAP`. Using computer programs, check the results obtained in Problem 16.4.

[1] `https://www.gap-system.org`

17 Singular homology

A **singular simplex** [1] is a continuous map from a standard simplex to a topological space. More precisely, for $n \geq 0$ an n-dimensional singular simplex in a topological space X is a continuous map $\sigma : \Delta_n \to X$, where Δ_n is the standard n-simplex.

Let $S_n(X)$ be the free abelian group generated by all n-dimensional singular simplices in X. As a set

$$S_n(X) = \left\{ \sum_{i=1}^{m} n_i \sigma_i \mid \sigma_i : \Delta_n \to X,\ n_i \in \mathbb{Z},\ m \in \mathbb{Z}^+ \right\}.$$

Each element of $S_n(X)$ is a finite sum of integer multiples of n-dimensional singular simplices, called a **singular n-chain**.

For $n \geq 1$ consider an n-dimensional singular simplex

$$\begin{aligned} \sigma : \Delta_n &\to X \\ (t_0, t_1, \ldots, t_n) &\mapsto \sigma(t_0, t_1, \ldots, t_n). \end{aligned}$$

For $0 \leq i \leq n$ define the **ith face** of σ to be the $(n-1)$-dimensional singular simplex

$$\begin{aligned} \partial_n^i(\sigma) : \Delta_{n-1} &\to X \\ (t_0, t_1, \ldots, t_{n-1}) &\mapsto \sigma(t_0, t_1, \ldots, t_{i-1}, 0, t_i, \ldots, t_{n-1}). \end{aligned}$$

Define the **boundary** of σ to be the $(n-1)$-dimensional singular chain

$$\partial_n(\sigma) = \sum_{i=0}^{n} (-1)^i \partial_n^i(\sigma).$$

Intuitively one can think of $\partial_n(\sigma)$ as $\sigma|_{\partial \Delta_n}$.

The boundary map ∂_n is extended linearly to become a group homomorphism from $S_n(X)$ to $S_{n-1}(X)$. We let $\partial_0 = 0$.

Example. A 0-dimensional singular simplex in X can be identified with a point in X.

A 1-dimensional singular simplex is a continuous map $\sigma(t_0, t_1)$ with $t_0, t_1 \in [0, 1]$ and $t_0 + t_1 = 1$. Its image is a curve between the points $A = \sigma(1, 0)$ and $B = \sigma(0, 1)$. Its boundary is $-A + B$.

A 2-dimensional singular simplex is a continuous map $\sigma(t_0, t_1, t_2)$ with $t_0, t_1, t_2 \in [0, 1]$ and $t_0 + t_1 + t_2 = 1$. Its image is a "curved triangle" between the points $A = \sigma(1, 0, 0)$, $B = \sigma(0, 1, 0)$, and $C = (0, 0, 1)$. Intuitively, the image

[1] đơn hình suy biến

of the face ∂^0 is the "curved edge" BC, the image of ∂^1 is AC, and the image of ∂^2 is AB. The boundary is $\partial^0 - \partial^1 + \partial^2$. Intuitively, it is $BC - AC + AB$.

Similar to the case of simplicial complex 16.3, we have:

17.1 Proposition. $\partial_{n-1} \circ \partial_n = 0$, $\forall n \geq 2$.

Proof. Let σ be a singular n-simplex. From definition:

$$\begin{aligned}
(\partial_{n-1}\partial_n)(\sigma) &= \partial_{n-1}\left(\sum_{i=0}^{n}(-1)^i \partial_n^i(\sigma)\right) \\
&= \sum_{j=0}^{n-1}(-1)^j \partial_{n-1}^j \left(\sum_{i=0}^{n}(-1)^i \partial_n^i(\sigma)\right) \\
&= \sum_{0\leq i\leq n, 0\leq j\leq n-1} (-1)^{i+j}\partial_{n-1}^j \partial_n^i(\sigma) \\
&= \sum_{0\leq j<i\leq n} (-1)^{i+j}\partial_{n-1}^j \partial_n^i(\sigma) + \\
&\quad + \sum_{0\leq i\leq j\leq n-1} (-1)^{i+j}\partial_{n-1}^j \partial_n^i(\sigma) \\
&= \sum_{0\leq j<i\leq n} (-1)^{i+j}\partial_{n-1}^j \partial_n^i(\sigma) + \\
&\quad + \sum_{0\leq i<k\leq n,\ (k=j+1)} (-1)^{i+k-1}\partial_{n-1}^{k-1} \partial_n^i(\sigma) \\
&= \sum_{0\leq j<i\leq n} (-1)^{i+j}\partial_{n-1}^j \partial_n^i(\sigma) + \\
&\quad + \sum_{0\leq j<i\leq n} -(-1)^{i+j}\partial_{n-1}^{i-1} \partial_n^j(\sigma).
\end{aligned}$$

For $0 \leq j < i \leq n$:

$$\begin{aligned}
\partial_{n-1}^j \partial_n^i(\sigma)(t_0, \dots, t_{n-2}) &= \partial_n^i \sigma(t_0, \dots, t_{j-1}, 0, t_j, \dots, t_{n-2}) \\
&= \sigma(t_0, \dots, t_{j-1}, 0, t_j, \dots, t_{i-2}, 0, t_{i-1}, \dots, t_{n-2}),
\end{aligned}$$

while

$$\begin{aligned}
\partial_{n-1}^{i-1} \partial_n^j(\sigma)(t_0, \dots, t_{n-2}) &= \partial_n^j \sigma(t_0, \dots, t_{i-2}, 0, t_{i-1}, \dots, t_{n-2}) \\
&= \sigma(t_0, \dots, t_{j-1}, 0, t_j, \dots, t_{i-2}, 0, t_{i-1}, \dots, t_{n-2}),
\end{aligned}$$

thus $\partial_{n-1}^j \partial_n^i = \partial_{n-1}^{i-1} \partial_n^j$. So $\partial_{n-1} \circ \partial_n = 0$. □

As in simplicial homology, elements of $\ker(\partial_n)$ are often called (singular) **closed chains** or **cycles**, while elements of $\mathrm{Im}(\partial_{n+1})$ are called **exact chains** or **boundaries**. Observe that ***exact chains are closed chains***, ***boundaries are cycles***.

17.2 Example. For $n \geq 1$, suppose $\sigma_n : \Delta_n \to X$, $\sigma_n(\Delta_n) = \{x_0\}$ is a constant map, a constant n-dimensional singular simplex. For each $0 \leq i \leq n$ the ith-face of σ is the map $\partial^i \sigma_n = \sigma_{n-1} : \Delta_{n-1} \to X$, $\sigma_{n-1}(\Delta_{n-1}) = \{x_0\}$. Thus

$$\partial \sigma_n = \sum_{i=0}^{n} (-1)^i \partial^i \sigma = \begin{cases} 0, & n \text{ is odd} \\ \sigma_{n-1}, & n \text{ is even.} \end{cases}$$

So except dimensional 0, a constant singular simplex is a cycle only in odd dimensions, in which case it is a boundary, the boundary of a constant singular simplex.

Definition. *For $n \geq 0$ the n-dimensional* **singular homology group** *of a topological space X is defined to be the quotient group*

$$H_n(X) = \frac{\ker(\partial_n)}{\operatorname{Im}(\partial_{n+1})}.$$

Two n-dimensional chains belong to the same homology class, in which case they are said to be **homologous** *, if and only if their difference is the boundary of an $(n+1)$-dimensional chain.*

Example. From Example 17.2, constant odd dimensional singular simplices are null-homologous (i.e representing the 0 class in the homology groups).

Example. Denoting by $\{\text{pt}\}$ a space containing only one point, then we can work out that $H_n(\{\text{pt}\}) = 0$ for $n \geq 1$ and $H_0(\{\text{pt}\}) = \langle \text{pt} \rangle \cong \mathbb{Z}$.

Proposition. *If X is path-connected then $H_0(X) \simeq \mathbb{Z}$, generated by any point of X. In general $H_0(X)$ is generated by one point in each path-connected components of X.*

Proof. Let x_0 and x_1 be two points in X. A continuous path from x_0 to x_1 gives rise to a singular 1-simplex $\sigma : \Delta_1 \to X$ such that $\sigma(0,1) = x_0$ and $\sigma(1,0) = x_1$. The boundary of this singular simplex is $\partial\sigma = x_1 - x_0 \in \operatorname{Im} \partial_1$. Thus $[x_0] = [x_1] \in H_0(X) = S_0(X)/\operatorname{Im} \partial_1$. $\square$

Induced homomorphism

Let X and Y be topological spaces and let $f : X \to Y$ be continuous. For any n-singular simplex σ let $f_\#(\sigma) = f \circ \sigma$, then extend $f_\#$ linearly, we get a group homomorphism $f_\# : S_n(X) \to S_n(Y)$.

Example. $\text{Id}_* = \text{Id}$.

Lemma. *$\partial \circ f_\# = f_\# \circ \partial$. As a consequence $f_\#$ brings cycles to cycles, boundaries to boundaries.*

Proof. Because both $f_\#$ and ∂ are linear we only need to prove $\partial(f_\#(\sigma)) = f_\#(\partial(\sigma)$ for any n-singular simplex $\sigma : \Delta_n \to X$. We have

$$f_\#(\partial(\sigma)) = f_\#\left(\sum_{i=0}^{n}(-1)^i\partial^i(\sigma)\right) = \sum_{i=0}^{n}(-1)^i f\circ\partial^i(\sigma).$$

On the other side: $\partial(f_\#(\sigma)) = \sum_{i=0}^{n}(-1)^i\partial^i(f\circ\sigma)$. Notice that

$$\begin{aligned}\partial^i(f\circ\sigma)(t_0,\dots,t_{n-1}) &= (f\circ\sigma)(t_0,\dots,t_{i-1},0,t_i,\dots,t_{n-1})\\ &= f(\sigma(t_0,\dots,t_{i-1},0,t_i,\dots,t_{n-1}))\\ &= f(\partial^i\sigma(t_0,\dots,t_{n-1})).\end{aligned}$$

Thus $\partial^i(f\circ\sigma) = f\circ\partial^i(\sigma)$. From this the result follows. ◻

As a consequence $f_\#$ induces a group homomorphism:

$$\begin{aligned} f_* : H_n(X) &\to H_n(Y).\\ [c] &\mapsto [f_\#(c)].\end{aligned}$$

Lemma. $(g\circ f)_* = g_*\circ f_*$.

Proof. Since the maps involved are linear it is sufficient to check this property on each n-singular simplex σ:

$$\begin{aligned}(g\circ f)_*([\sigma]) = [(g\circ f)_\#(\sigma)] = [(g\circ f)\circ\sigma] = [g\circ(f\circ(\sigma))] = g_*([f\circ\sigma]) =\\ = g_*(f_*([\sigma])).\end{aligned}$$

◻

A simple application of the above lemma gives us an important result:

Theorem (topological invariance of homology). *If $f : X \to Y$ is a homeomorphism then $f_* : H_n(X) \to H_n(Y)$ is an isomorphism.*

Proof. Apply the above lemma to the pair f and f^{-1}. ◻

Thus ***singular homology is a topological invariant***.

Relation between homology and homotopy

A path is a continuous map from an interval, while a singular simplex is a continuous map from a standard line segment. We can anticipate a relation between them.

Take the following linear parametrization of the standard 1-simplex:

$$\begin{aligned}\varphi : [0,1] &\to \Delta_1 \\ t &\mapsto (1-t, t).\end{aligned}$$

A path $\gamma : [0,1] \to X$ is then associated with a singular 1-simplex $\gamma \circ \varphi^{-1} : \Delta_1 \to X$. In this way we can identify a path with the corresponding singular simplex.

We have the following result:

17.3 Proposition (homotopic paths are homologous). *If the paths α and β are path-homotopic then the corresponding singular simplices $\alpha \circ \varphi^{-1}$ and $\beta \circ \varphi^{-1}$ are homologous.*

Proof. Let F be a path-homotopy from α to β, so $F : [0,1] \times [0,1] \to X$ is continuous such that $F(0,t) = F(1,t) =$ constant, $F(s,0) = \alpha(s)$, $F(s,1) = \beta(s)$. This proof is based on the following simple idea, see Fig. 17.4. Denote $O = (0,0)$, $A = (1,0)$, $B = (1,1)$, $C = (0,1)$. The restrictions $F|_{OAC}$ and $F|_{ACB}$ are "more or less" singular 2-simplices. Their boundaries are "more or less" $\partial(F|_{OAC}) = F|_{AC} - F|_{CO} + F|_{OA}$ and $\partial(F|_{ACB}) = F|_{CB} - F|_{AB} + F|_{AC}$. Notice that $F|_{OA} = \alpha$, $F|_{CB} = \beta$, while $F|_{AB}$ and $F|_{CO}$ are "more or less" constant simplices hence are null-homologous by Example 17.2. Thus

$$\partial(F|_{OAC} - F|_{ACB}) = \alpha - \beta.$$

So $\alpha - \beta$ is the boundary of a chain, hence is null-homologous.

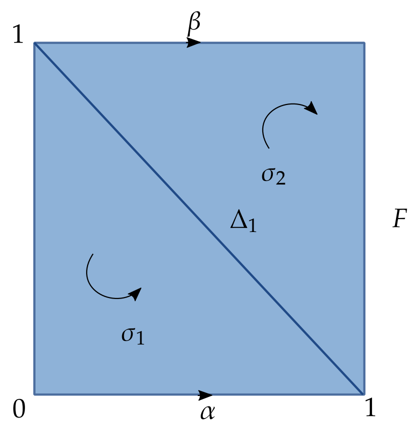


Figure 17.4: The idea that $\partial(\sigma_1 - \sigma_2) = \alpha - \beta$.

The term "more or less" above means up to a reparametrization, so we can write more accurately as follows. Let $\psi_1 : \Delta_2 \to OAC$ be the map $\psi_1(t_0, t_1, t_2) = t_0 O + t_1 A + t_2 C = (t_1, t_2)$, and let $\psi_2 : \Delta_2 \to ACB$ be the map $\psi_2(t_0, t_1, t_2) = t_0 A + t_1 C + t_2 B = (t_0 + t_2, t_1 + t_2)$. Let $\sigma_1 = F \circ \psi_1$ and $\sigma_2 = F \circ \psi_2$ then σ_1 and σ_2 are truly singular simplices. We can work out that for $(t_0, t_1) \in \Delta_1 = AC$,

$$\partial^0 \sigma_1(t_0, t_1) = \sigma_1(0, t_0, t_1) = F(\psi_1(0, t_0, t_1)) = F(t_0, t_1),$$

thus $\partial^0 \sigma_1 = F|_{\Delta_1}$. Similarly we find $\partial^1 \sigma_1(t_0, t_1) = F(0, t_1) =$ constant, and $\partial^2 \sigma_1(t_0, t_1) = F(t_1, 0) = \alpha(t_1) = \alpha \circ \varphi^{-1}(t_0, t_1)$. Thus $\partial \sigma_1$ is homologous to $F|_{\Delta_1} + \alpha \circ \varphi^{-1}$. In the same way we find that $\partial^0 \sigma_2(t_0, t_1) = F(t_1, 1) = \beta(t_1) = \beta \circ \varphi^{-1}(t_0, t_1)$, $\partial^1 \sigma_2(t_0, t_1) = F(1, t_1) =$ constant, $\partial^2 \sigma_2(t_0, t_1) = F(t_0, t_1)$, thus $\partial \sigma_2$ is homologous to $\beta \circ \varphi^{-1} + F|_{\Delta_1}$. Hence $\partial(\sigma_1 - \sigma_2)$ is homologous to $\alpha \circ \varphi^{-1} - \beta \circ \varphi^{-1}$, and so $\alpha \circ \varphi^{-1}$ and $\beta \circ \varphi^{-1}$ are homologous.

For a different approach, see Problem 17.13. □

If γ is a loop at x_0, then $\gamma \circ \varphi^{-1}$ is a singular 1-simplex which is a cycle, since

$$\partial(\gamma \circ \varphi^{-1}) = \gamma \circ \varphi^{-1}(0, 1) - \gamma \circ \varphi^{-1}(1, 0) = \gamma(1) - \gamma(0) = x_0 - x_0 = 0.$$

We can now consider the map

$$\begin{aligned} \pi_1(X, x_0) &\longrightarrow H_1(X) \\ [\gamma] &\mapsto [\gamma \circ \varphi^{-1}]. \end{aligned}$$

Proposition 17.3 implies that this map is well-defined. This map is often called the ***Hurewicz map***.

Proposition. *The Hurewicz map is a group homomorphism.*

Proof. Let α and β be loops at x_0. We need to show that $(\alpha \cdot \beta) \circ \varphi^{-1}$ is homologous to $\alpha \circ \varphi^{-1} + \beta \circ \varphi^{-1}$. For this we can use the idea in Fig. 17.5. We

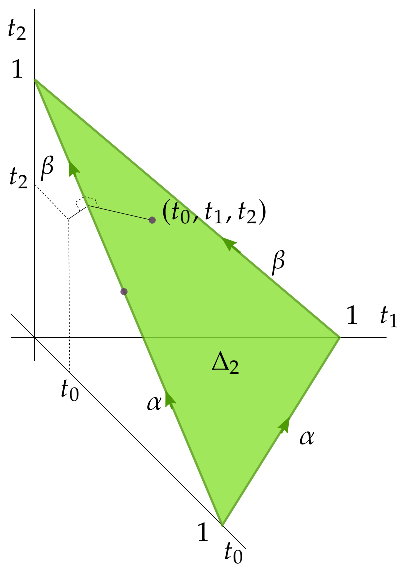


Figure 17.5: Idea for proving that the Hurewics map is a homomorphism.

project each point (t_0, t_1, t_2) orthogonally to the edge of Δ_2 opposite to the vertex $(0, 1, 0)$. We can work out that the projection is at $\left(\frac{t_0-t_2+1}{2}, 0, \frac{t_2-t_0-1}{2}\right)$. We define a singular 2-simplex $\sigma : \Delta_2 \longrightarrow X$ as

$$\sigma(t_0, t_1, t_2) = \begin{cases} \alpha(t_2 - t_0 + 1), & t_2 \leq t_0, \\ \beta(t_2 - t_0), & t_2 \geq t_0. \end{cases}$$

We can check that, as suggested in Fig. 17.5:

$$\partial\sigma = \beta \circ \varphi^{-1} - (\alpha \cdot \beta) \circ \varphi^{-1} + \alpha \circ \varphi^{-1}.$$

This implies $\alpha \circ \varphi^{-1} + \beta \circ \varphi^{-1} - (\alpha \cdot \beta) \circ \varphi^{-1}$ is homologous to 0, in other words, $(\alpha \cdot \beta) \circ \varphi^{-1}$ is homologous to $\alpha \circ \varphi^{-1} + \beta \circ \varphi^{-1}$. □

Since H_1 is commutative, any element of $\pi_1(X, x_0)$ of the form $a \cdot b \cdot a^{-1} \cdot b^{-1}$, called a commutator, is sent by the Hurewicz homomorphism to 0.

Theorem (Hurewicz theorem). *Let X be path-connected. The Hurewicz homomorphism is surjective with kernel the subgroup generated by all commutators. Thus* ***the first homology group $H_1(X)$ is isomorphic to the abelianization of the fundamental group $\pi_1(X)$****.*

Proofs are available in [Vic94, p. 108], [Hat01, p. 166].

Example. For the sphere S^n, since $\pi_1(S^n) = 1$ if $n > 1$ and $\pi_1(S^1) \cong \mathbb{Z}$, in both cases are abelian groups, we get $H_1(S^n) = 0$ if $n > 1$ and $H_1(S^1) \cong \mathbb{Z}$.

Example. The fundamental group of the torus was found to be

$$\pi_1(T_1) \cong \langle a, b \mid ab = ba \rangle .$$

It is already abelian, so it is isomorphic to $H_1(T_1)$.

Example. The fundamental group of the Klein bottle was found in 15.9 to be

$$\pi_1(K) \cong \left\langle a, b \mid aba^{-1}b \right\rangle .$$

By Hurewicz theorem the first homology group is

$$H_1(K) \cong \left\langle a, b \mid aba^{-1}b = 1, ab = ba \right\rangle = \left\langle a, b \mid b^2 = 1, ab = ba \right\rangle \cong \mathbb{Z} \oplus \mathbb{Z}_2.$$

In 15.5 we have the fundamental groups of compact surfaces. Using Hurewicz theorem, we can find the first homology groups of compact surfaces, this is Problem 17.14.

The following result says that ***homology is a homotopy invariant*** [Vic94, p. 13]:

Theorem (homotopy invariance of homology). *If $f : X \to Y$ is a homotopy equivalence then $f_* : H_n(X) \to H_n(Y)$ is an isomorphism.*

Mayer-Vietoris sequence

Theorem. *Let X be a topological space. Suppose $U, V \subset X$ and $int(U) \cup int(V) = X$. Then there is an exact chain complex, called the **Mayer-Vietoris sequence**:*

$$\cdots \to H_n(U \cap V) \xrightarrow{(i_*, j_*)} H_n(U) \oplus H_n(V) \xrightarrow{\psi_*} H_n(U \cup V) \xrightarrow{\Delta} H_{n-1}(U \cap V) \to \cdots$$
$$\cdots \to H_0(U \cup V) \to 0.$$

Here i and j are the inclusion maps from $U \cap V$ to U and V respectively.

In the Mayer-Vietoris sequence above the maps ψ_* and Δ are certain homomorphisms which can be written in more details. A proof is beyond the scope of this note [Vic94, p. 22].

The Mayer-Vietoris sequence allows us to study the homology of a space from homologies of subspaces, in a similar manner to the Van Kampen theorem.

We now use the Mayer-Vietoris sequence to compute the homology of the sphere:

Theorem. *For $m \geq 1$,*

$$H_n(S^m) \cong \begin{cases} \mathbb{Z}, & \text{if } n = 0, m \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let U and V be the upper hemisphere and the lower hemisphere slightly enlarged, for example, $U = S^m \setminus \{(0, \ldots, 0, -1)\}$ and $V = S^m \setminus \{(0, \ldots, 0, 1)\}$. The Mayer-Vietoris sequence for this pair gives an exact short sequence:

$$H_n(U) \oplus H_n(V) \to H_n(S^m) \to H_{n-1}(U \cap V) \to H_{n-1}(U) \oplus H_n(V).$$

Notice that for $m \geq 1$, U and V are contractible, while $U \cap V$ is homotopic to S^{m-1}. We get for $n \geq 2$ a short exact sequence

$$0 \to H_n(S^m) \to H_{n-1}(S^{m-1}) \to 0.$$

This implies $H_n(S^m) \cong H_{n-1}(S^{m-1})$ for $n \geq 2$. The problem now reduces to computation of $H_1(S^m)$.

Consider the exact sequence:

$$H_1(U) \oplus H_1(V) \to H_1(S^m) \xrightarrow{\Delta} H_0(U \cap V) \xrightarrow{(i_*, j_*)} H_0(U) \oplus H_0(V).$$

Since $H_1(U) = H_1(V) = 0$, it follows that Δ is injective.

For $m \geq 2$ a point $x \in U \cap V$ generates $H_0(U \cap V)$ as well as $H_0(U)$ and $H_0(V)$. Therefore the maps i_* and j_* are injective. This implies $\mathrm{Im}(\Delta) = 0$. This can happen only when $H_1(S^m) = 0$.

For $m = 1$ the intersection $U \cap V$ has two path-connected components. Let x and y be points in each connected component. If $mx + ny \in H_0(U \cap V)$ then $i_*(mx+ny) = j_*(mx+ny) = mx+nx = (m+n)x$. Thus $\ker(i_*, j_*) = \{mx-my = m(x-y) \mid m \in \mathbb{Z}\}$. This implies $H_1(S^1) \cong \mathrm{Im}(\Delta) = \ker(i_*, j_*) = \langle x-y \rangle \cong \mathbb{Z}$. □

17.6 Corollary. *For $n \geq 2$ there cannot be any retraction from the disk D^n to its boundary S^{n-1}.*

The proof is similar to the 2-dimensional case (14.5).

Proof. Suppose there is a retraction $r : D^n \to S^{n-1}$. Let $i : S^{n-1} \hookrightarrow D^n$ be the inclusion map. From the diagram $S^{n-1} \xrightarrow{i} D^n \xrightarrow{r} S^{n-1}$ we have $r \circ i = \mathrm{id}_{S^{n-1}}$, therefore on the $(n-1)$-dimensional homology groups $(r \circ i)_* = r_* \circ i_* = \mathrm{id}_{H_{n-1}(S^{n-1})}$, implying that $r_* : H_{n-1}(D^n) \to H_{n-1}(S^{n-1})$ is onto. But this is not possible for $n \geq 2$, since $H_{n-1}(D^n)$ is trivial while $H_{n-1}(S^{n-1})$ is not. □

A proof of this result in differentiable setting is presented in 24.5.

Just as in the case of dimension two (14.6), the Brouwer fixed point theorem follows, with the same proof:

17.7 Theorem (Brouwer fixed point theorem). *Any continuous map from the disk D^n to itself has a fixed point.*

By 17.22 and 17.21 we get a more general version:

Theorem (generalized Brouwer fixed point theorem). *Any continuous map from a compact convex subset of a Euclidean space to itself has a fixed point.*

For more on singular homology one can read [Vic94].

Problems

17.8. Prove 17.1.

17.9. Compute the homology groups of $S^2 \times [0,1]$.

17.10. Compute the fundamental group and the homology groups of the Euclidean space $\mathbb{R}^3$ minus a straight line.

17.11. Compute the fundamental group and the homology groups of the Euclidean space $\mathbb{R}^3$ minus two intersecting straight lines.

17.12. Compute the fundamental group and the homology groups of $\mathbb{R}^3 \setminus S^1$.

17.13. This is a sketch of a different approach to prove 17.3. From a path-homotopy F from α to β we can construct a singular simplex σ such that $\partial^0\sigma = \alpha\circ\varphi^{-1}$, $\partial^1\sigma = \beta\circ\varphi^{-1}$, and $\partial^2\sigma$ = constant, which proves that $\alpha\circ\varphi^{-1} - \beta\circ\varphi^{-1}$ is a boundary.

Let $A = (1, \frac{1}{2})$, $B = (0, 1)$. Let $\psi : \Delta_2 \to [0,1]\times[0,1]$ given by $\psi(t_0, t_1, t_2) = t_0B + t_1O + t_2A$. In this way we bring Δ_2 to the triangle BOA. We define $\tilde{\sigma}$ on BOA such that on OA it is α and on AB it is β. A way to this is to expand BOA vertically to $[0,1]\times[0,1]$. Namely, let

$$\tilde{\sigma}(x,y) = \begin{cases} F\left(x, \frac{1}{2} + \frac{y-\frac{1}{2}}{1-x}\right), & 0 \le x < 1, \frac{1}{2}x \le y \le -\frac{1}{2}x + 1,\\ F\left(1, \frac{1}{2}\right), & x = 1, y = \frac{1}{2}. \end{cases}$$

Check that $\tilde{\sigma}$ is continuous. Let $\sigma = \tilde{\sigma}\circ\psi$. Check that σ has the desired properties.

17.14. ✓ Using Hurewicz theorem, find the first homology groups of compact surfaces from their fundamental groups in 15.5.

17.15. Show that if X has k connected components X_i, $1 \le i \le k$, then $S_n(X) \cong \bigoplus_{i=1}^k S_n(X_i)$ and $H_n(X) \cong \bigoplus_{i=1}^k H_n(X_i)$.

17.16. Show that if A and B are open, $A \cap B$ is contractible then $H_i(A\cup B) \cong H_i(A) \oplus H_i(B)$ for $i \ge 2$. Is this true if $i = 0, 1$?

17.17. Using Problem 17.16, compute the homology groups of $S^2 \vee S^4$.

17.18. Compute the fundamental group and the homology groups of $S^2 \cup \{(0,0,z) \mid -1 \le z \le 1\}$.

17.19. Let U be an open subset of $\mathbb{R}^2$ and $a \in U$. Show that

$$H_1(U \setminus \{a\}) \cong H_1(U) \oplus H_1(\mathbb{R}^2 \setminus \{a\}).$$

17.20. Let A be a retract of X via a retraction $r : X \to A$. Let $i : A \hookrightarrow X$ be the inclusion map. Show that the induced homomorphism $i_* : H_n(A) \to H_n(X)$ is injective while $r_* : H_n(X) \to H_n(A)$ is surjective.

17.21. Show that if every continuous map from X to itself has a fixed point and Y is homeomorphic to X then every continuous map from Y to itself has a fixed point.

17.22. * Show that:

(a) Any convex compact subset of $\mathbb{R}^n$ with empty interior is homeomorphic to the closed interval D^1.

(b) Any convex compact subset of $\mathbb{R}^n$ with non-empty interior is homeomorphic to the disk D^n.

(c) The boundary of any convex compact subset of $\mathbb{R}^n$ with non-empty interior is homeomorphic to the sphere S^{n-1}.

For related questions, see 10.19 and 3.33.

17.23. Is the Brouwer fixed point theorem correct for open balls? for spheres? for tori?

17.24. Let A be an $n \times n$ matrix whose entries are all non-negative real numbers. We will derive the Frobenius theorem which says that A must have a real non-negative eigenvalue.

(a) Suppose that A is not singular. Check that the map $v \mapsto \frac{Av}{\|Av\|}$ brings $Q = \{(x_1, x_2, \ldots, x_n) \in S^{n-1} \mid x_i \geq 0,\ 1 \leq i \leq n\}$ to itself.

(b) Prove that Q is homeomorphic to the closed ball D^{n-1}.

(c) Use the continuous Brouwer fixed point theorem to prove that A has a real non-negative eigenvalue.

18 Homology of cell complexes

In this section we consider homology for a special class of topological spaces: cell complexes, including simplicial complexes.

Degrees of maps on spheres

A continuous map $f : S^n \to S^n$ induces a homomorphism $f_* : H_n(S^n) \to H_n(S^n)$. We know $H_n(S^n) \cong \mathbb{Z}$, so there is a generator a (the other generator is $-a$) such that $H_n(S^n) = \langle a \rangle$. Choosing such a generator is called giving S^n a homological orientation. Then $f_*(a) = ma$ for a certain integer m, called the **topological degree** of f, denoted by $\deg f$.

Example. If f is the identity map then $\deg f = 1$. If f is the constant map then $\deg f = 0$.

If f and g are maps from S^n to itself then because $(g \circ f)_* = g_* \circ f_*$ we have $\deg(g \circ f) = \deg(g) \cdot \deg(f)$.

Example. Let $f : S^1 \to S^1$, given by $f((\cos 2\pi t, \sin 2\pi t)) = (\cos 2\pi nt, \sin 2\pi nt)$, or by using complex number notation, $f(z) = z^n$. Then $\deg(f) = n$. This is convincing but a detail explanation needs some works, see Problem 18.1.

Relative homology groups

Let A be a subspace of X. Viewing each singular simplex in A as a singular simplex in X, we have a natural inclusion $S_n(A) \hookrightarrow S_n(X)$. In this way $S_n(A)$ is a normal subgroup of $S_n(X)$. The boundary map ∂_n induces a homomorphism $\partial_n : S_n(X)/S_n(A) \to S_{n-1}(X)/S_{n-1}(A)$, giving a chain complex

$$\cdots \to S_n(X)/S_n(A) \xrightarrow{\partial_n} S_{n-1}(X)/S_{n-1}(A) \xrightarrow{\partial_{n-1}} S_{n-2}(X)/S_{n-2}(A) \to \cdots$$

The homology groups of this chain complex is called the **relative homology groups** of the pair (X, A), denoted by $H_n(X, A)$.

If $f : X \to Y$ is continuous and $f(A) \subset B$ then as before it induces a homomorphism $f_* : H_n(X, A) \to H_n(Y, B)$.

Homology of cell complexes

Let X be a (finite) cellular complex. Recall that X^n denotes the n-dimensional skeleton of X. Suppose that X^n is obtained from X^{n-1} by attaching the n-dimensional disks $D_1^n, D_2^n, \ldots, D_{c_n}^n$. Let $e_1^n, e_2^n, \ldots, e_{c_n}^n$ be the corresponding cells. Then

$$H_n(X^n, X^{n-1}) \cong \langle e_1^n, e_2^n, \ldots, e_{c_n}^n \rangle = \{\sum_{i=1}^{c_n} m_i e_i^n \mid m_i \in \mathbb{Z}\}.$$

This is probably intuitively convincing, a proof is available in [Hat01, p. 137]. If $c_n = 0$ then let the group be 0.

Consider the sequence

$$C(X) = \cdots \xrightarrow{d_{n+1}} H_n(X^n, X^{n-1}) \xrightarrow{d_n} H_{n-1}(X^{n-1}, X^{n-2}) \xrightarrow{d_{n-1}} \cdots$$
$$\cdots \xrightarrow{d_2} H_1(X^1, X^0) \xrightarrow{d_1} H_0(X).$$

Here the map d_n is given by

$$d_n(e_i^n) = \sum_{j=1}^{c_{n-1}} d_{i,j} e_j^{n-1},$$

where the integer number $d_{i,j}$ is the degree of the following map on spheres:

$$S_i^{n-1} = \partial D_i^n \to X^{n-1} \to X^{n-1}/X^{n-2} = S_1^{n-1} \vee S_2^{n-1} \vee \cdots \vee S_{c_{n-1}}^{n-1} \to S_j^{n-1}.$$

Of course if $c_{n-1} = 0$ then $d_n = 0$.

The following is the main tool for computing homology of cellular complexes:

Theorem. *The sequence $C(X)$ is a chain complex and its homology coincides with the singular homology of X.*

A proof is outside the scope of this note, see [Hat01, p. 139].

As an application we get:

Theorem (homology groups of surfaces). *The homology groups of a connected compact orientable surface S of genus $g \geq 0$ is*

$$H_n(S) \cong \begin{cases} \mathbb{Z}, & \text{if } n = 0, 2 \\ \mathbb{Z}^g & \text{if } n = 1. \end{cases}$$

For more on cellular homology one can read [Hat01, p. 137].

Problems

18.1. Show in detail that if $f : S^1 \to S^1$, given by $f((\cos 2\pi t, \sin 2\pi t)) = (\cos 2\pi n t, \sin 2\pi n t)$, or by using complex number notation, $f(z) = z^n$, then $\deg(f) = n$.

18.2. Using cellular homology compute the homology groups of the following spaces:

(a) The Klein bottle.

(b) $S^1 \vee S^1$.

(c) $S^1 \vee S^2$.

(d) $S^2 \vee S^3$.

18.3. Show that two compact connected surfaces are homeomorphic if and only if their homologies are isomorphic.

Other topics

For course projects and further study, besides many topics already suggested in previous sections, below there are several more topics.

Invariance of dimension

That the Euclidean spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ are not homeomorphic is not easy. It is a consequence of the following difficult theorem of L. Brouwer in 1912:

18.4 Theorem (Invariance of domain). *If two subsets of the Euclidean $\mathbb{R}^n$ are homeomorphic and one set is open then the other is also open.*

This theorem is often proved using Algebraic Topology, see for instance [Mun00, p. 381], [Vic94, p. 34], [Hat01, p. 126].

Corollary. *The Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homeomorphic if $m \neq n$.*

Proof. Suppose that $m < n$. We can check that the inclusion map $\mathbb{R}^m \to \mathbb{R}^n$ given by $(x_1, x_2, \dots, x_m) \mapsto (x_1, x_2, \dots, x_m, 0, \dots, 0)$ is a homeomorphism onto its image $A \subset \mathbb{R}^n$. Since $\mathbb{R}^n$ is open in $\mathbb{R}^n$, if A is homeomorphic to $\mathbb{R}^n$ then by Invariance of dimension A is open in $\mathbb{R}^n$. But A is clearly not open in $\mathbb{R}^n$. □

Jordan curve theorem

The following is an important result of plane topology:

Theorem (Jordan curve theorem). *A simple, continuous, closed curve separates the plane into two disconnected regions. More concisely, if C is a subset of the Euclidean plane homeomorphic to the circle then $\mathbb{R}^2 \setminus C$ has two connected components.*

Nowadays proofs of this theorem is often found in texts in Algebraic Topology, [Mun00, p. 390].

The Poincaré conjecture

Theorem (Poincaré conjecture). *A compact topological n-dimensional manifold which is homotopic to the n-dimensional sphere is homeomorphic to the n-dimensional sphere.*

The proof of this statement is the result of a cumulative effort of many mathematicians, including Stephen Smale (for dimension ≥ 5, early 1960s), Michael Freedman (for dimension 4, early 1980s), and Grigory Perelman (for dimension 3, early 2000s). For dimension 2 it is elementary, see Problem 15.16.

More topics

- Cayley graphs of groups, [Hat01, p. 77, 78].
- Delaunay triangulation (may restrict to dimension 2), important in Computational Topology, [EH10, p. 80].
- π_1 as a functor in Category Theory, [Peter May, *A concise course in Algebraic Topology*, p. 13–15].

Part III Differential Topology

Differential Topology studies topology in the smooth category, meaning that the maps considered are smooth. The main objects are smooth manifolds, generalizations of curves and surfaces.

This course focuses on discussing important and representative examples, main ideas of notions, arguments and proofs.

The course benefits students interested in theoretical mathematics, especially in areas related to geometry and physics, and is useful for areas such as differential equations, nonlinear analysis, partial differential equations, mechanics, ….

For attending this course, taking a prior course in General Topology is not required, however is strongly recommended.

Some proofs in this part are more difficult. The readers may choose to focus on attempting to understand the main ideas of the arguments, assuming some results.

Along the way we take opportunities to study again parts of mathematics such as linear algebra, analysis, and differential equations, as needed.

We closely follow John Milnor's lectures in [Mil97]. There are several textbooks partly suitable to undergraduate students such as [GP74], [Sas11], [Dun18], [Tu13]. There are recent freely available lecture notes [Qui], [Kui20].

For more advanced treatments, the book [Hir76] can serve as a reference for some advanced topics, the textbook [Lee13] supplies many detailed proofs, the book [DFN85] contains integrated presentation of both topology and geometry.

Parts of this subject are often discussed in courses and books on differential manifolds.

19 Smooth manifolds

We introduced topological (continuous) manifolds in Section 10, the readers who have not previously encountered the idea of manifold can read an introduction there.

Roughly, manifolds are objects for which local descriptions are available. A smooth manifold is a space which can be locally described by smooth functions. A smooth manifold is said to be locally diffeomorphic to $\mathbb{R}^n$. It is possible to bring the differential and integral calculus from $\mathbb{R}^n$ to smooth manifolds, providing a setting for generalizations of vector calculus of curves and surfaces to higher dimensions.

For simplicity we consider only smooth manifolds in Euclidean spaces instead of abstract smooth manifolds.

Smooth maps on open subsets of $\mathbb{R}^n$

We summarize here several results about derivatives of functions defined on open sets in $\mathbb{R}^n$, see for instance [Spi65] or [Lan97] for more details.

In this part we always assume that $\mathbb{R}^n$ has the Euclidean topology.

Let $D \subset \mathbb{R}^k$ and let x be an ***interior point*** of D. The partial derivative of the function $f : D \to \mathbb{R}$ with respect to the ith variable at x is defined to be the real number

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h\to 0} \frac{f(x + he_i) - f(x)}{h},$$

where $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ form the canonical vector basis for $\mathbb{R}^n$. This is the rate of change of the value of f compared to the value of its ith variable. The **gradient vector** of f is

$$\operatorname{grad} f(x) = \nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right).$$

Consider $f : D \to \mathbb{R}^l$. The matrix of all first order partial derivatives of the component functions of f is called the **Jacobian matrix** of f, denoted by $J_f(x)$, Jf_x, or $J(f)(x)$:

$$J_f(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{1\le i\le l,\ 1\le j\le k}.$$

If there is a linear map $f'(x) : \mathbb{R}^k \to \mathbb{R}^l$ such that there is a ball $B(x, \epsilon) \subset D$ and a function $r : B(x, \epsilon) \to \mathbb{R}^l$ satisfying:

$$f(x + h) = f(x) + f'(x)(h) + r(h),\ \forall h \in B(x, \epsilon)$$

and $\lim_{h\to 0} \frac{r(h)}{|h|} = 0$, then $f'(x)$ is called the (Fréchet) **derivative** of f at x. The derivative is a linear approximation of the function: $f(x+h) \approx f(x)+f'(x)(h)$. In particular if the function is linear then its derivative is itself. We will also denoted $f'(x)$ by $df(x)$ or df_x. It should be emphasized that the derivative is a linear map, not a number or a matrix.

We can deduce that

$$f'(x)(h) = \lim_{t\to 0} \frac{f(x+th)-f(x)}{t}.$$

If $\|h\| = 1$ then $f'(x)(h)$ is the directional derivative of f at x in the direction of h, measuring rate of change of f in the direction of h at x. In particular $f'(x)(e_i) = \frac{\partial f}{\partial x_i}(x)$.

If f has continuous partial derivatives at x then f has derivative at x, and the linear map $f'(x)$ can be represented in the canonical linear bases by the Jacobian matrix $J_f(x)$, i.e. $f'(x)(h) = J_f(x)\cdot h$.

Let U, V, W be open subsets of $\mathbb{R}^k$, $\mathbb{R}^l$, $\mathbb{R}^p$ respectively, let $f : U \to V$ and $g : V \to W$ be smooth maps, we have the chain rule

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

We can take higher order partial derivatives. For example, $\frac{\partial^2 f}{\partial x_i \partial x_j}$ is a partial derivative of order 2. If f is continuous and all partial derivatives of all orders of f exist and are continuous at x then we say that f is **smooth** at x. In common notation in Analysis, it means $f \in C^\infty = \bigcap_{k=0}^{\infty} C^k$. The reader may notice that the properties above hold for C^1 functions. In several places in this part we need C^2 functions. For simplicity in this part we assume that our functions are smooth (to any order).

Smooth manifolds

Previously smoothness of a function is defined only for interior points of the domain. Since a manifold in a Euclidean space such as a curve in the plane or a surface in $\mathbb{R}^3$ can have empty interior, we may not be able to discuss smoothness at all. So first we extend the notion of smoothness to points not in the interior of the domain.

Definition. *Let $D \subset \mathbb{R}^k$ and let $x \in D$ be a boundary point of D. Then $f : D \to \mathbb{R}^l$ is said to be* **smooth** *at x if f can be extended to a function which is smooth (in the previous sense) in an open neighborhood in $\mathbb{R}^k$ of x. Precisely, f is smooth at x if there is an open set $U \subset \mathbb{R}^k$ containing x and function $F : U \to \mathbb{R}^l$ such that F is smooth at every point of U and $F|_{U\cap D} = f|_{U\cap D}$.*

If f is smooth at every point of D then we say that f is smooth on D, or $f \in C^\infty(D)$.

Proposition. *Compositions of smooth functions are smooth.*

Proof. Let $f : D \to E \subset \mathbb{R}^l$ and let $g : E \to \mathbb{R}^p$. Suppose that f is smooth at $x \in D$ and g is smooth at $f(x)$. There is an extension of f to smooth $F : U \to \mathbb{R}^l$ where U is an open neighborhood of x in $\mathbb{R}^k$, and there is an extension of g to smooth $G : V \to \mathbb{R}^p$ where V is an open neighborhood of $f(x)$ in $\mathbb{R}^l$. Since F is continuous, $U' = F^{-1}(V)$ is open in U, hence is open in $\mathbb{R}^k$, and contains x. Then the composition $G \circ F : U' \to \mathbb{R}^p$ is smooth, therefore $g \circ f$ is smooth at x. □

Notice that we do not define derivatives for functions which are smooth in this extended sense, since there is no obvious way to do it yet: different extensions may have different derivatives (compare 19.14).

Let $X \subset \mathbb{R}^k$ and $Y \subset \mathbb{R}^l$. Then $f : X \to Y$ is a **diffeomorphism** [1] if it is bijective and both f and f^{-1} are smooth. If there is a diffeomorphism from X to Y then we say that X and Y are **diffeomorphic**.

If a function is smooth at a point, it must be continuous at that point, therefore a diffeomorphism is a homeomorphism.

Example. Any two balls in $\mathbb{R}^n$ are diffeomorphic. Comparing with 3.3, given $B(a,r)$ and $B(b,s)$ we can take the diffeomorphism $\varphi : B(a,r) \to B(b,s)$, $\varphi(x) = \frac{s}{r}(x-a)+b$.

Further, any open ball in $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$. Comparing with 3.3, we may consider the function $\varphi : B(0,1) \to \mathbb{R}^n$, $\varphi(x) = \frac{1}{\sqrt{1-\|x\|^2}}x$. The inverse map is $\varphi^{-1}(y) = \frac{1}{\sqrt{1+\|y\|^2}}y$. Both maps are smooth.

Definition. *A subspace $M \subset \mathbb{R}^k$ is a* **smooth manifold** *of dimension $m \in \mathbb{Z}^+$ if every point in M has a neighborhood in M which is diffeomorphic to $\mathbb{R}^m$.*

Sometimes for convenience we also talk about **0-dimensional manifold**. *We shall use this convention: a 0-dimensional manifold M in $\mathbb{R}^k$, $k \geq 1$, is a discrete subspace of $\mathbb{R}^k$. Every point x of this manifold has a neighborhood in $\mathbb{R}^k$ whose intersection with M consists only x.*

A diffeomorphism is a homeomorphism, therefore a smooth manifold is a topological manifold, see Section 10. In differential topology unless stated otherwise manifolds mean smooth manifolds.

The following observation, although seems to be less intuitive than our original definition, is technically more convenient to use:

Proposition. *A subspace $M \subset \mathbb{R}^k$ is a smooth manifold of dimension m if every point in M has an open neighborhood in M which is diffeomorphic to an open subset of $\mathbb{R}^m$.*

[1] vi đồng phôi, vi phôi

Proof. The proof is similar to the case of topological manifold in Section 10. Let M be a manifold. Suppose that U is a neighborhood of x in M and $\phi : U \to \mathbb{R}^m$ is a diffeomorphism. There is an open subset U' of M such that $x \in U' \subset U$. Since ϕ is a homeomorphism on U and U' is open in U, the set $\phi(U')$ is open in $\mathbb{R}^m$. Take a ball $B(\phi(x), r) \subset \phi(U')$. Since ϕ is continuous on U', the set $U'' = \phi^{-1}(B(\phi(x), r))$ contains x and is open in U' hence is open in M. The restriction $\phi|_{U''} : U'' \to B(\phi(x), r)$ is a diffeomorphism. We have just shown that any point in the manifold has an open neighborhood diffeomorphic to an open ball in $\mathbb{R}^m$.

For the reverse direction, suppose that U is an open neighborhood of x in M and $\phi : U \to V$ is a diffeomorphism where V is open in $\mathbb{R}^m$. Take a ball $B(\phi(x), r) \subset V$ and let $U' = \phi^{-1}(B(\phi(x), r)$ then U' contains x and is an open set in U hence is open in M. The restriction $\phi|_{U'} : U' \to B(\phi(x), r)$ is a diffeomorphism. Recall that any open ball in $\mathbb{R}^m$ is diffeomorphic to $\mathbb{R}^m$, hence the ball $B(\phi(x), r)$ is diffeomorphic to $\mathbb{R}^m$ via a diffeomorphism ψ. Then $\psi \circ \phi|_{U'}$ is a diffeomorphism from U' to $\mathbb{R}^m$. □

By this result, each point x in a smooth manifold has an open neighborhood U in M and a diffeomorphism $\varphi : U \to V$ where V is an open subset of $\mathbb{R}^m$. The pair (V, φ^{-1}) is called a **local parametrization** at x. The pair (U, φ) is called a **local coordinate** or a **chart** [1] at x. The set of pairs (U, φ) at all points is called an **atlas** [2] for M. We can think of a manifold as a space which has an atlas, each page of that atlas is a chart mapping a region in the space.

If M and N are two smooth manifolds in $\mathbb{R}^k$ and $M \subset N$ then we say that M is a **submanifold** of N.

Example. Any open subset of $\mathbb{R}^m$ is a smooth manifold of dimension m.

19.1 Proposition. *The graph of a smooth function $f : D \to \mathbb{R}^l$, where $D \subset \mathbb{R}^k$ is an open set, is a smooth manifold of dimension k.*

Proof. Let $G = \{(x, f(x)) \mid x \in D\} \subset \mathbb{R}^{k+l}$ be the graph of f. The map $x \mapsto (x, f(x))$ from D to G is smooth. Its inverse is the projection $(x, y) \mapsto x$. This projection is the restriction of the projection given by the same formula from $\mathbb{R}^{k+l}$ to $\mathbb{R}^k$, which is a smooth map. Therefore D is diffeomorphic to G. □

Example (curves and surfaces). The graph of a smooth function $y = f(x)$ for $x \in (a, b)$ is a 1-dimensional smooth manifold in $\mathbb{R}^2$, often called a smooth curve.

The graph of a smooth function $z = f(x, y)$ for $(x, y) \in D$, where D is an open set in $\mathbb{R}^2$ is a 2-dimensional smooth manifold in $\mathbb{R}^3$, often called a smooth surface.

[1] tờ bản đồ
[2] tập bản đồ

Example **(circle).** Let $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. It is covered by four neighborhoods which are half circles, corresponding to points $(x, y) \in S^1$ such that $x > 0$, $x < 0$, $y > 0$ and $y < 0$. Each of these neighborhoods is diffeomorphic to $(-1, 1)$. For example consider the projection $U = \{(x, y) \in S^1 \mid x > 0\} \longrightarrow (-1, 1)$ given by $(x, y) \mapsto y$. The map $(x, y) \mapsto y$ is smooth on $\mathbb{R}^2$, so it is smooth on U. The inverse map $y \mapsto (\sqrt{1 - y^2}, y)$ is smooth on $(-1, 1)$. Therefore the projection is a diffeomorphism.

Remark. $\mathbb{R}^m$ cannot be diffeomorphic to $\mathbb{R}^n$ if $m \neq n$. Indeed, if $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is a diffeomorphism then the derivative $df_x : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is a linear isomorphism of vector spaces, and so $m = n$. As a corollary, a non-empty smooth manifold has a unique dimension.

The Inverse function theorem

Recall the following important result in Analysis (see for instance [Spi65]):

Theorem **(Inverse function theorem).** *Let $f : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ be smooth. If $\det(Jf_x) \neq 0$ then there is an open neighborhood U of x and an open neighborhood V of $f(x)$ such that $f|_U \colon U \longrightarrow V$ is a diffeomorphism.*

Briefly, if df_x is an isomorphism then f is a diffeomorphism locally at x.

The result is usually stated for continuously differentiable function (i.e. C^1), but the result for smooth functions follows, since the Jacobian matrix of the inverse map is the inverse matrix of the Jacobian of the original map, and the entries of an inverse matrix can be obtained from the entries of the original matrix via smooth operations, namely

$$A^{-1} = \frac{1}{\det A} A^*, \qquad A^*_{i,j} = (-1)^{i+j} \det(A^{j,i}) \tag{19.2}$$

where $A^{j,i}$ is obtained from A by omitting the jth row and ith column [Lan87, p. 177].

Smooth invariance of domain is a simple corollary (compare the continuous case 18.4):

19.3 Corollary **(smooth invariance of domain).** *If two subsets of the Euclidean $\mathbb{R}^k$ are diffeomorphic and one set is open then the other set is also open.*

Proof. Let $U, V \subset \mathbb{R}^k$ and let U be open. Let $f : U \longrightarrow V$ be a diffeomorphism. Let $y \in V$. There is $x \in U$ such that $f(x) = y$. Since f^{-1} is smooth at y, there is an open subset W of $\mathbb{R}^k$ containing y and a smooth function $g : W \longrightarrow \mathbb{R}^k$ such that $g|_{W \cap V} = f^{-1}$. Let $U' = f^{-1}(W \cap V)$, then U' is open in U, hence in $\mathbb{R}^k$. We have $g \circ (f|_{U'}) = \mathrm{id}_{U'}$. Taking derivative at x we get $dg_y \circ df_x = \mathrm{id}_{\mathbb{R}^k}$. This implies df_x is an isomorphism. By the Inverse function theorem, there is an open set V' of $\mathbb{R}^k$ containing y on which f^{-1} is defined. So $V' \subset V$. Thus V is open. $\square$

19.4 Proposition. *A subspace M of $\mathbb{R}^k$ is an m-dimensional smooth manifold if and only if for each $x_0 \in M$ there is an open set U in $\mathbb{R}^m$ and a map $\varphi : U \to M$ such that*

(a) $\varphi(U)$ is an open neighborhood of x_0,

(b) $\varphi : U \to \varphi(U)$ is a homeomorphism,

(c) φ is smooth,

(d) $d\varphi(u) : \mathbb{R}^m \to \mathbb{R}^k$ is injective for each $u \in U$.

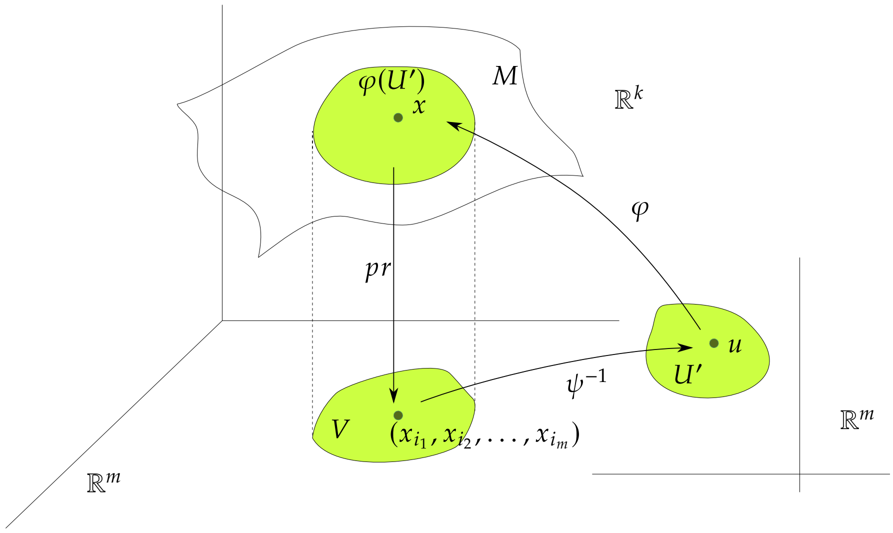
That $d\varphi(u)$ is injective can be stated alternatively as rank $d\varphi(u) = m$, where the rank of a linear map between two linear spaces is the dimension of the image of the map. We observe that in this statement it is not required to check that the inverse map φ^{-1} is smooth. The statement is used as the definition of smooth manifold in [Mun91, p. 196], also see [Spi65, p. 111]. When $m = 2$ and $k = 3$ this is the definition of regular surfaces in [dC76, p. 52]. The reader can also compare this with the notions of regular curves and regular surfaces used in Multivariable Calculus [Vgt3].

Proof. The necessary direction: Since φ^{-1} is smooth at x_0 it has an extension to a function F which is smooth in an open neighborhood O of x_0 in $\mathbb{R}^k$ and coincides with φ^{-1} in $O \cap M$. We can take O small enough that $O \cap M \subset \varphi(U)$. Since $F \circ \varphi$ is the identity function in a neighborhood $\varphi^{-1}(O \cap M)$ of u, taking derivative we get $dF(\varphi(u)) \circ d\varphi(u) = \mathrm{id}_{\mathbb{R}^m}$, thus $d\varphi(u)$ is injective.

The sufficient direction: The main content of the statement is that under the assumptions it is possible to deduce that φ^{-1} is smooth. Since $d\varphi(u)$ is injective, the Jacobian matrix $J\varphi(u)$ is of rank m, having m independent rows corresponding to m components $x_{i_1}, x_{i_2}, \dots, x_{i_m}$. Thus the $m \times m$-matrix

$$\begin{pmatrix} \frac{\partial x_{i_1}}{\partial u}(u) & \frac{\partial x_{i_2}}{\partial u}(u) & \cdots & \frac{\partial x_{i_m}}{\partial u}(u) \end{pmatrix}$$

is non-singular. Denote the correspondence $u \mapsto \big(x_{i_1}(u), x_{i_2}(u), \dots, x_{i_m}(u)\big)$ by ψ. Apply the Inverse function theorem to ψ, there is an open neighborhood U' of u in $\mathbb{R}^m$ and an open neighborhood V of $\big(x_{i_1}(u), x_{i_2}(u), \dots, x_{i_m}(u)\big)$ in $\mathbb{R}^m$ such that ψ is a diffeomorphism from U' onto V. Replacing U' by $U' \cap U$ if necessary, we can assume $U' \subset U$.



Let pr be the projection $(x_1, x_2, \dots, x_k) \mapsto (x_{i_1}, x_{i_2}, \dots, x_{i_m})$. The following composition

$$\begin{array}{ccccc} \varphi(U') & \xrightarrow{pr} & V & \xrightarrow{\psi^{-1}} & U' \\ (x_1, x_2, \dots, x_k) & \mapsto & (x_{i_1}, x_{i_2}, \dots, x_{i_m}) & \mapsto & u \end{array}$$

gives $\varphi^{-1} = \psi^{-1} \circ pr$ on $\varphi(U')$. Since φ is a homeomorphism, $\varphi(U')$ is an open neighborhood of x in M. Since $\psi^{-1} \circ pr$ is smooth on $pr^{-1}(V)$ which is an open subset of $\mathbb{R}^k$, its restriction to $\varphi(U')$ is smooth at x. Thus φ^{-1} is smooth at x.

Notice that we do need to use the assumption that φ is a homeomorphism so that $\varphi(U')$ is an open neighborhood of x in M, otherwise the argument may not work, see Problem 19.11. $\square$

Example. The graph of the function $y = |x|$ is not a smooth 1-dimensional manifold in $\mathbb{R}^2$. Indeed, if the graph, denoted by G, is a smooth 1-manifold then there is a parametrization $\varphi(t) = (x(t), y(t))$ of a neighborhood of $\varphi(0) = (0, 0)$ on G. Since $\varphi'(0) = (x'(0), y'(0)) \neq 0$, either $x'(0) \neq 0$ or $y'(0) \neq 0$. Repeating the above argument, we find that either there is a neighborhood of $(0, 0)$ in G which is the graph of a smooth function on x, or there is a neighborhood of $(0, 0)$ in G which is the graph of a smooth function on y. Obviously the second possibility cannot happen. For the first possibility, the smooth function must be $y = |x|$, but this function is not smooth.

Problems

19.5. Show that if X and Y are diffeomorphic and X is an m-dimensional manifold then so is Y.

19.6. Show that any open subset of a smooth manifold is a smooth manifold.

19.7. Show that the sphere $S^n = \{(x_1, x_2, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\}$ is a smooth manifold of dimension n, covered by the hemispheres.

Prove that in another way, by using two stereographic projections (3.8), one from the North Pole and one from the South Pole.

19.8. Show that the hyperboloid $x^2 + y^2 - z^2 = 1$ is a manifold.

19.9. The torus can be obtained by rotating around the z axis a circle on the xOz plane not intersecting the z axis (see 8.10). Show that the torus is a smooth manifold.

19.10. Consider the union of the curve $y = x^3 \sin \frac{1}{x}$, $x \neq 0$ and the point $(0, 0)$. Is it a smooth manifold?

19.11. Is the trace of the path $\gamma(t) = (\sin(2t), \sin(t))$, $t \in (0, 2\pi)$ (the figure 8) a smooth manifold?

19.12. Consider the union of the curve $y = \sin \frac{1}{x}$, $x > 0$ and the segment $\{(0, y) | -1 \leq y \leq 1\}$ (the Topologist's sine curve, see Section 4.9). Is it a smooth manifold?

19.13. Is the surface $x^2 + y^2 - z^2 = 0$ a manifold?

19.14. From our definition, a smooth function f defined on $D \subset \mathbb{R}^k$ does not necessarily have partial derivatives defined at boundary points of D. However, show that if D is the closure of an open subspace of $\mathbb{R}^k$ then the partial derivatives of f are defined and are continuous on D. For example, $f : [a, b] \to \mathbb{R}$ is smooth if and only if f has right-derivative at a and left-derivative at b, or equivalently, f is smooth on an open interval (c, d) containing $[a, b]$.

19.15. A simple closed regular path is a map $\gamma : [a, b] \to \mathbb{R}^m$ such that γ is injective on $[a, b)$, γ is smooth, $\gamma^{(k)}(a) = \gamma^{(k)}(b)$ for all integer $k \geq 0$, and $\gamma'(t) \neq 0$ for all $t \in [a, b]$. Show that the trace of a simple closed regular path is a smooth 1-dimensional manifold.

19.16. Show that the trefoil knot (see 8.40) is a smooth 1-dimensional manifold (in fact it is diffeomorphic to the circle S^1, but this is more difficult).

19.17. ✓ Show that a connected smooth manifold is also path-connected.

19.18. Show that on a path-connected smooth manifold any two points can be connected by a piecewise smooth path.

19.19. ✓ Let M be a smooth manifold and let φ be a local parametrization of a neighborhood of $x \in M$. Show that a map $f : M \to \mathbb{R}^l$ is smooth at x if and only if $f \circ \varphi$ is smooth at $\varphi^{-1}(u)$.

20 Tangent spaces and derivatives

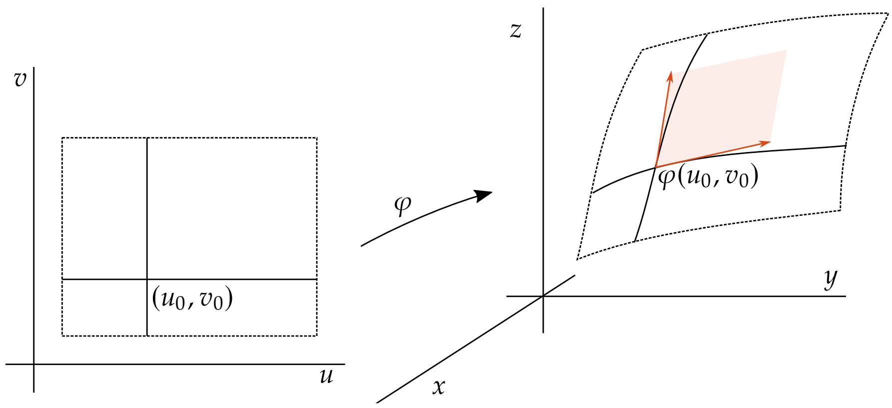
In Calculus we use tangent lines to approximate curves, and tangent planes to approximate surfaces. The tangent lines and tangent planes are linear spaces, simpler than the curves and the surfaces, while giving effective local approximations.

While living on the surface of the Earth, we often think about a neighborhood on the ground as a region of a flat plane. Now we know that is not correct – the ground is not flat – but still the approximation is often good enough in a "small" neighborhood.

Now we develop this idea of local linear approximation to higher dimensions.

Tangent spaces of manifolds

To motivate the definition of tangent spaces of manifolds we recall the notion of tangent spaces of surfaces. Consider a parametrized surface in $\mathbb{R}^3$ given by $\varphi(u, v) = (x(u, v), y(u, v), z(u, v))$. Consider a point $\varphi(u_0, v_0)$ on the surface. Near to (u_0, v_0) if we fix $v = v_0$ and only allow u to change then we get a parametrized path $\varphi(u, v_0)$ passing through $\varphi(u_0, v_0)$. The velocity vector of the path $\varphi(u, v_0)$ is a "tangent vector" to the path at the point $\varphi(u_0, v_0)$, and is given by the partial derivative with respect to u, that is, $\frac{\partial \varphi}{\partial u}(u_0, v_0)$. Similarly we have another "tangent vector" $\frac{\partial \varphi}{\partial v}(u_0, v_0)$. Then the "tangent space" of the surface at $\varphi(u_0, v_0)$ is the plane spanned by the above two tangent vectors.



We can think of a manifold as a high dimensional surface, our definition of tangent space of manifold is a natural generalization.

Definition. *Let M be an m-dimensional manifold in $\mathbb{R}^k$, let $x \in M$, and let φ be a parametrization of a neighborhood of x, and $x = \varphi(u)$. We define the* **tangent space** *of M at x, denoted by TM_x, or T_xM, or $T(M)(x)$, to be the vector subspace of $\mathbb{R}^k$*

spanned by the vectors $\frac{\partial\varphi}{\partial u_i}(u)$, $1 \leq i \leq m$. Each element of the tangent space is called a **tangent vector**.

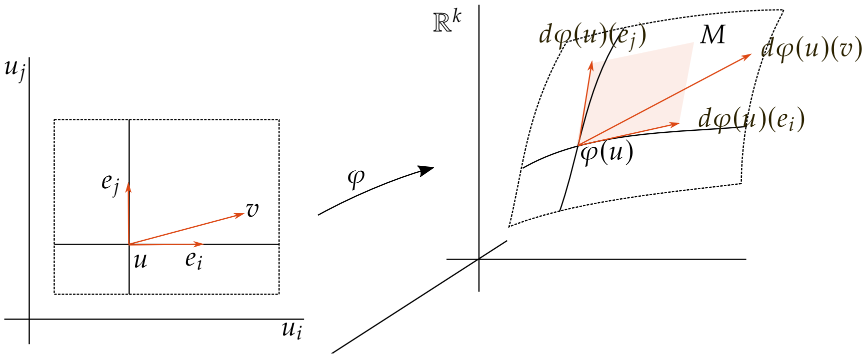


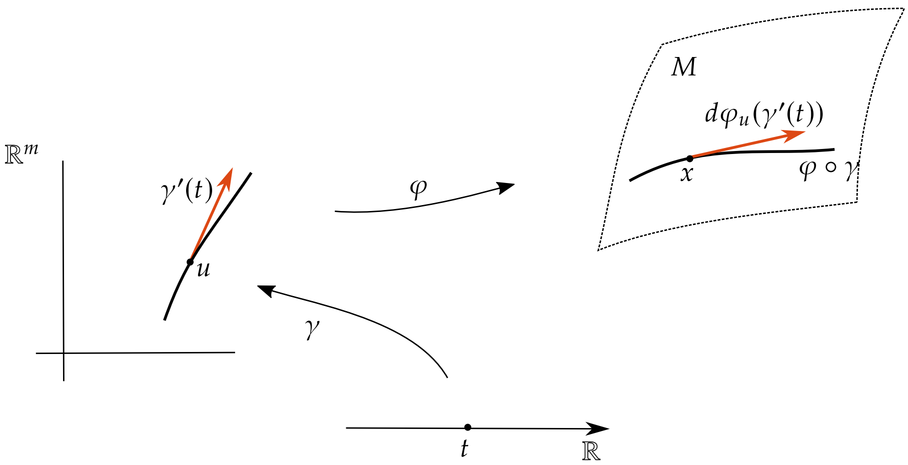
Figure 20.1: $d\varphi_u(v) = d\varphi_u\left(\sum_{i=1}^m v_ie_i\right) = \sum_{i=1}^m v_i d\varphi_u(e_i)$.

Recall that $\frac{\partial\varphi}{\partial u_i}(u) = d\varphi(u)(e_i)$. If v is a vector in $\mathbb{R}^m$, $v = \sum_{i=1}^m v_ie_i$, where $v_i \in \mathbb{R}$, then

$$d\varphi_u(v) = d\varphi_u\left(\sum_{i=1}^m v_ie_i\right) = \sum_{i=1}^m v_i d\varphi_u(e_i) = \sum_{i=1}^m v_i \frac{\partial\varphi}{\partial u_i}(u).$$

We deduce that $TM_x = d\varphi_u(\mathbb{R}^m)$. This means the derivative of the local parametrization brings the Euclidean space onto the tangent space.

There is another interpretation of tangent vector in terms of velocity vectors. A smooth path in $\mathbb{R}^m$ is a smooth map γ from an open interval of $\mathbb{R}$ to $\mathbb{R}^m$. The vector $\gamma'(t) \in \mathbb{R}^k$ is called the velocity vector of the path at t. The local parametrization φ then brings this path to $\varphi \circ \gamma$, a smooth path on M. The velocity vector of $\varphi \circ \gamma$ is $\frac{d}{dt}(\varphi \circ \gamma)(t) = d\varphi(u)(\gamma'(t))$, which is a tangent vector of M. Conversely, it can be seen easily that every tangent vector of M at x is the velocity vector of a smooth path in M at x (Problem 20.6).



Example. If M is an open set of $\mathbb{R}^k$, then M is a k-dimensional manifold, with the identity map as a parametrization, thus for each $x \in M$ the tangent space is $TM_x = \mathbb{R}^k$.

Example (curve). Consider a curve $y = f(x)$ in $\mathbb{R}^2$. Take the parametrization $\varphi(x) = (x, f(x))$. The tangent space at $(a, f(a))$ consists of the multiples of the vector $\varphi'(a) = (1, f'(a))$, thus it is given by the equation $y = f'(a)x$. Notice the difference with the equation given in Calculus, $y = f(a) + f'(a)(x - a)$, here the tangent space is a linear space containing 0, while in Calculus a line does not need to be a linear space and does not need to contain 0.

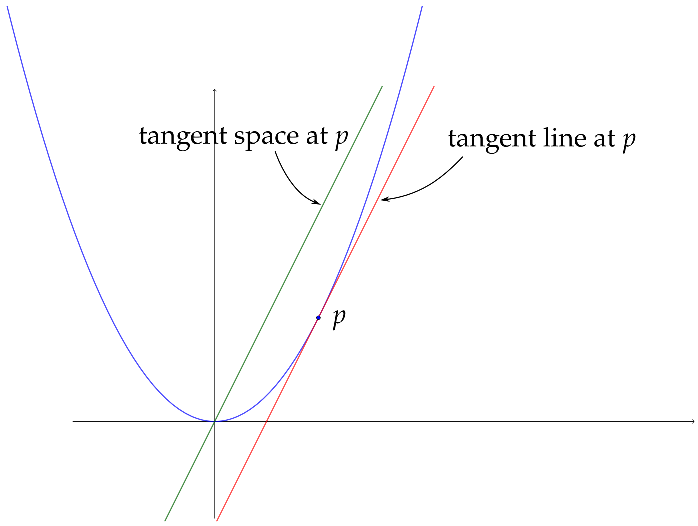


Figure 20.2: The tangent space of a manifold is a linear space, as in the subject of Linear Algebra. Here a tangent vector, as in Linear Algebra, has no initial point. A tangent line or a tangent plane in in the subject of Calculus can be a translation of the tangent space. In figures, for demonstration we usually draw tangent lines, tangent planes, and tangent vectors passing through tangent points.

Example (surface). Consider a surface $z = f(x, y)$ in $\mathbb{R}^3$. Take the parametrization $\varphi(x, y) = (x, y, f(x, y))$. The tangent space at $(a, b, f(a, b))$ consists of the linear combinations of the vectors $\varphi_x(a, b) = (1, 0, f_x(a, b))$ and $\varphi_y(a, b) = (0, 1, f_y(a, b))$, thus it is given by the equation $z = xf_x(a, b) + yf_y(a, b)$. Similarly to the case of curves, notice the difference with the equation for the tangent plane given in Calculus, $z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$.

Example (circle). Consider the circle S^1. At a point $(x, y) \neq (1, 0)$ there is a parametrization $(x, y) = \varphi(t) = (\cos t, \sin t)$. The derivative of φ brings any real number v to the vector $d\varphi(t)(v) = (\cos t, \sin t)'(v) = v(-\sin t, \cos t)$. Thus the tangent line at $(\cos t, \sin t)$ is spanned by the vector $(-\sin t, \cos t)$. Notice that $(-\sin t, \cos t) \cdot (\cos t, \sin t) = 0$, thus this tangent line is perpendicular to the vector (x, y).

In another approach, let us use velocity vectors. Let $(x(t), y(t))$ be any path on S^1. The tangent space of S^1 at (x, y) is spanned by the velocity vector

$(x'(t), y'(t))$ if this vector is not 0. Since $x(t)^2 + y(t)^2 = 1$, differentiating both sides with respect to t we get $x(t)x'(t) + y(t)y'(t) = 0$, or in other words $(x'(t), y'(t))$ is perpendicular to $(x(t), y(t))$. Thus the tangent space is the line perpendicular to the radius.

Lemma. *The tangent space does not depend on the choice of parametrization.*

Proof. Consider the following diagram, where U, U' are open, φ and φ' are parametrizations of open neighborhood of $x \in M$.

$$\begin{tikzcd} & M & \\ U \arrow[ur, "\varphi"] \arrow[rr, "\varphi'^{-1}\circ\varphi"'] & & U' \arrow[ul, "\varphi'"'] \end{tikzcd}$$

Note that the map $\varphi'^{-1} \circ \varphi$ is to be understood as follows. We have that $\varphi(U) \cap \varphi'(U')$ is a neighborhood of $x \in M$. Restricting to $\varphi^{-1}(\varphi(U) \cap \varphi'(U'))$, the map $\varphi'^{-1} \circ \varphi$ is well-defined, and is a diffeomorphism. The above diagram gives us, with any $v \in \mathbb{R}^m$:

$$d\varphi_u(v) = d\varphi'_{\varphi'^{-1}\circ\varphi(u)}\big(d(\varphi'^{-1} \circ \varphi)_u(v)\big).$$

Thus any tangent vector with respect to the parametrization φ is also a tangent vector with respect to the parametrization φ'. We conclude that the tangent space does not depend on the choice of parametrization. □

The above statement can also be seen from the interpretation of tangent vectors as velocity vectors of smooth paths.

Proposition. *If M is an m-dimensional manifold then the tangent space TM_x is an m-dimensional linear space.*

Proof. Given a parametrization φ, by 19.4 the map $d\varphi_u$ is an injective linear map, thus the image of $d\varphi_u$ is an m-dimensional vector space. □

Derivatives of maps on manifolds

Here we develop the Fréchet derivative in Calculus to manifold. The derivative at a point is a linear map, from a linear space to a linear space, giving rate of change of the value of the function when the variable changes in a direction. The linear spaces should be the tangent spaces.

Let $M \subset \mathbb{R}^k$ and $N \subset \mathbb{R}^l$ be manifolds of dimensions m and n respectively. Let $f : M \to N$ be smooth. Let $x \in M$. There is a neighborhood W of x in $\mathbb{R}^k$ and a smooth extension F of f to W. The derivative of f is defined to be the restriction of the derivative of F. Precisely:

Definition. *The* **derivative** *of f at x is defined to be $df_x = dF_x|_{TM_x}$, i.e.*

$$\begin{array}{rcl} df_x : TM_x & \to & TN_{f(x)} \\ h & \mapsto & df_x(h) = dF_x(h). \end{array}$$

We need to show that the derivative is well-defined:

Lemma. *$df_x(h) \in TN_{f(x)}$ and does not depend on the choice of F.*

Thus, although as noted in the previous section a smooth map defined on a general subset of $\mathbb{R}^k$ may not have derivative, on a manifold the derivative can be defined.

Proof. Consider the following diagram

$$\begin{array}{ccc} W \cap M & \xrightarrow{\;F\;} & N \\ \varphi\uparrow & & \uparrow\psi \\ Uar[ru, ''f \circ \varphi''] & \xrightarrow[\psi^{-1}\circ f\circ\varphi]{} & V \end{array}$$

Let us explain this diagram. Take a parametrization ψ of a neighborhood of $f(x)$. Assume that $\varphi(u) = x$, $\psi(v) = f(x)$, $h = d\varphi_u(w)$.Then $f^{-1}(\psi(V))$ is an open neighborhood of x in M. We can find an open set W in $\mathbb{R}^k$ such that $(W \cap M) \subset f^{-1}(\psi(V))$ and $W \cap M$ is an open neighborhood of x in M parametrized by φ, and f has an extension to a smooth function F on W.

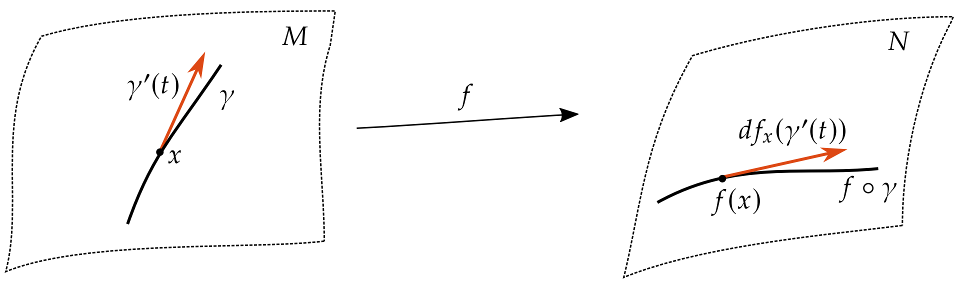
The diagram indicates two ways to write df_x. For any $w \in \mathbb{R}^m$:

$$df_x(d\varphi_u(w)) = dF_x\big(d\varphi_u(w)\big) = d(F \circ \varphi)_u(w) = d(f \circ \varphi)_u(w),$$

which shows that df_x does not depend on the choice of F, and furthermore

$$df_x(d\varphi_u(w)) = d(f\circ\varphi)_u(w) = d(\psi\circ(\psi^{-1}\circ f\circ\varphi))_u(w) = d\psi_v\big(d(\psi^{-1}\circ f\circ\varphi)_u(w)\big),$$

which shows that df_x has value in $TN_{f(x)}$. $\square$



The above statement can also be seen in the following way. A tangent vector at $x \in M$ is the velocity $\gamma'(t)$ of a path γ with $\gamma(t) = x$. The map f brings this path to N, giving the path $f \circ \gamma$. The velocity vector of this new path at $f(x)$ is

$$\left(f \circ \gamma\right)'(t) = (F \circ \gamma)'(t) = dF_x(\gamma'(t)) = df_x(\gamma'(t)).$$

Thus the derivative brings the velocity vector of a path in to the velocity of the corresponding path. This interpretation of df_x does not use an extension F, however it is not clear that df_x is linear.

Example. Let us consider simple cases:

(a) If M is an open set in $\mathbb{R}^k$ and $f : M \to N$ then df_x is the usual derivative of a smooth function in Calculus.

(b) If $\mathrm{Id} : M \to M$ is the identity map then $d\mathrm{Id}_x = \mathrm{Id}_{TM_x}$ for any $x \in M$.

(c) If $f : M \to N$ is a constant map then $df_x = 0$.

Proposition (chain rule). *If $f : M \to N$ and $g : N \to P$ are smooth maps between manifolds then*

$$d(g \circ f)_x = dg_{f(x)} \circ df_x.$$

In other words, the following commutative diagram

$$\begin{tikzcd} & N \arrow[dr, "g"] & \\ M \arrow[ur, "f"] \arrow[rr, "g\circ f"'] & & P \end{tikzcd}$$

induces the following commutative diagram

$$\begin{tikzcd} & \mathbb{R}^l \arrow[dr, "dg_y"] & \\ \mathbb{R}^k \arrow[ur, "df_x"] \arrow[rr, "d(g\circ f)_x"'] & & \mathbb{R}^p \end{tikzcd}$$

In mathematics, a diagram of maps among sets is often said to be a **commutative diagram** if going from one set to another set following composition of maps does not depend on the choice of a path.

Proof. There is an open neighborhood V of y in $\mathbb{R}^l$ and a smooth extension G of g to V. There is an open neighborhood U of x in $\mathbb{R}^k$ such that $U \subset F^{-1}(V)$ and there is a smooth extension F of f to U. Then $d(g \circ f)_x = d(G \circ F)_x|_{TM_x} = (dG_y \circ dF_x)|_{TM_x} = dG_y|_{TN_y} \circ dF_x|_{TM_x} = dg_y \circ df_x$. $\square$

Proposition. *If $f : M \to N$ is a diffeomorphism then $df_x : TM_x \to TN_{f(x)}$ is a linear isomorphism. In particular the dimensions of the two manifolds are same.*

Proof. Let $m = \dim M$ and $n = \dim N$. Since $df_x \circ df^{-1}_{f(x)} = \mathrm{Id}_{TN_{f(x)}}$ and $df^{-1}_{f(x)} \circ df_x = \mathrm{Id}_{TM_x}$ we deduce, via the rank of df_x, that $m \geq n$. Doing the same with $df^{-1}_{f(x)}$ we get $m \leq n$, hence $m = n$. Thus df_x must be a linear isomorphism. □

Problems

20.3. Calculate the tangent spaces of S^n.

20.4. Let V be an m-dimensional vector subspace of $\mathbb{R}^k$, let $p \in \mathbb{R}^k$, and let $M = p + V$, called an affine subspace of $\mathbb{R}^k$.

(a) Show that M is an m-dimensional smooth manifold.

(b) Find the tangent spaces of M.

(c) Let W be an n-dimensional vector subspace of $\mathbb{R}^l$, let $q \in \mathbb{R}^k$, and let $N = q + W$. Let $T : V \to W$ be a linear map. Define $\tilde{T} : M \to N$ by $\tilde{T}(p + v) = q + T(v)$. Find the derivatives of $\tilde{T}$.

20.5. Show that if M is a submanifold of N then TM_x is a linear subspace of TN_x.

20.6. Show that every tangent vector of M at x is the velocity vector at x of a smooth path in M going through x.

20.7 (Cartesian products of manifolds). Show that if $X \subset \mathbb{R}^k$ and $Y \subset \mathbb{R}^l$ are manifolds then $X \times Y \subset \mathbb{R}^{k+l}$ is also a manifold. Furthermore $T(X \times Y)_{(x,y)} = TX_x \times TY_y$. As an application, find the tangent space of the cylinder $S^1 \times \mathbb{R}$.

20.8. Calculate the derivative of the maps:

(a) $f : (0, 2\pi) \to S^1$, $f(t) = (\cos t, \sin t)$.

(b) $f : S^1 \to \mathbb{R}$, $f(x, y) = y$.

20.9. Let $M \subset \mathbb{R}^k$ be a closed subset and a smooth manifold and let $p \in \mathbb{R}^k$.

(a) Show that there exists $q \in M$ such that $d(p, q) = d(p, M) = \inf\{d(p, x) \mid x \in M\}$ (see 5.28).

(b) Show that $(p - q) \perp TM_q$ (in the Euclidean inner product of $\mathbb{R}^k$).

Compare this with a similar result on orthogonal projection often encountered in Functional Analysis [KF75], [TTV].

21 Regular values

The Implicit function theorem

Theorem (Implicit function theorem). *Suppose that*

$$\begin{aligned} f : \mathbb{R}^m \times \mathbb{R}^n &\to \mathbb{R}^n \\ (x, y) &\mapsto f(x, y) \end{aligned}$$

is smooth, $f(x_0, y_0) = 0$, and the $n \times n$ matrix $\frac{\partial f}{\partial y}(x_0, y_0)$ is non-singular, then there is a neighborhood $U \times V$ of (x_0, y_0) such that for each $x \in U$ there is a unique $g(x) \in V$ satisfying $f(x, g(x)) = 0$, furthermore the function g is smooth.

Proof. The main idea is to set $F(x, y) = (x, f(x, y))$ then apply the Inverse function theorem to F. We have

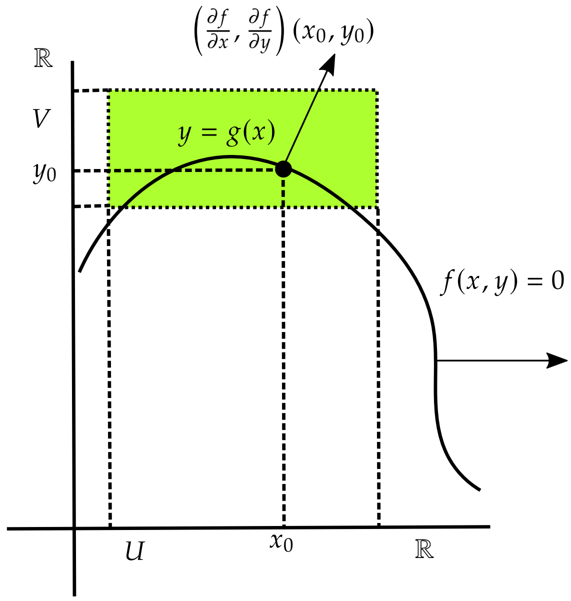
$$JF = \begin{pmatrix} I_{m\times m} & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix},$$

we thus $JF_{(x_0,y_0)}$ is non-singular. Applying the Inverse function theorem, there is an open neighborhood $W \times V$ of (x_0, y_0) such that $F_{|W\times V}$ is a diffeomorphism onto an open set $F(W \times V) \ni F(x_0, y_0) = (x_0, f(x_0, y_0)) = (x_0, 0)$. There is an open set $U \times V_1 \subset \mathbb{R}^m \times \mathbb{R}^n$ such that $(x_0, 0) \in U \times V_1 \subset F(W \times V)$. Now suppose $x \in U$. Since $(x, 0) \in U \times V_1$ there is a unique $(x_1, y) \in W \times V$ such that $F(x_1, y) = (x_1, f(x_1, y)) = (x, 0)$, thus we must have $x_1 = x$ and there is a unique $y \in V$ such that $f(x, y) = 0$. Let $g(x) = y$, then g is exactly the map

$$x \mapsto (x, 0) \overset{F^{-1}}{\mapsto} F^{-1}(x, 0) = (x, y) \mapsto y$$

which is smooth. □

The conclusion of the theorem implies that on $U \times V$ the implicit equation $f(x, y) = 0$ has a unique solution $y = g(x)$. It also implies that $\{(x, y) \in U \times V \mid f(x, y) = 0\} = \{(x, g(x)) \mid x \in U\}$. In other words, on the open neighborhood $U \times V$ of (x_0, y_0) the level set $f^{-1}(0)$ is the graph of the smooth function g.



21.1 Corollary. *Let $f : \mathbb{R}^{m+n} \longrightarrow \mathbb{R}^n$ be smooth and $f(x) = y$. If df_x is surjective then there is an open neighborhood of x in the level set $f^{-1}(y)$ which is diffeomorphic to an open subset of $\mathbb{R}^m$.*

Proof. All we need is a permutation of variables to bring the function to the form used by the Implicit function theorem. We can proceed as follows. Since df_x is onto, the Jacobian matrix $J(f)(x)$ is of rank n, therefore it has n independent columns, corresponding to n variables $x_{i_1}, x_{i_2}, \ldots, x_{i_n}$. Let σ be any permutation of $\{1, 2, \ldots, (m+n)\}$ such that $\sigma(i_1) = m+1, \sigma(i_2) = m+2, \ldots, \sigma(i_n) = m+n$. Let $\varphi : \mathbb{R}^{m+n} \longrightarrow \mathbb{R}^{m+n}$ be the map that permutes the variables as $(u_1, u_2, \ldots, u_{m+n}) \mapsto (u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(m+n)})$. Consider the function $f \circ \varphi$. By the chain rule

$$\frac{(f \circ \varphi)_i}{\partial u_j} = \sum_{k=1}^{m+n} \frac{\partial f_i}{\partial x_k} \cdot \frac{\partial x_k}{\partial u_j} = \sum_{k=1}^{m+n} \frac{\partial f_i}{\partial x_k} \cdot \frac{\partial u_{\sigma(k)}}{\partial u_j} = \frac{\partial f_i}{\partial x_{\sigma^{-1}(j)}}$$

For $j > m$ we get $\frac{(f \circ \varphi)_i}{\partial u_j} = \frac{\partial f_i}{\partial x_{i_{j-m}}}$, thus the last n columns of the matrix $J(f \circ \varphi))(u)$ are the above n independent columns of $J(f)(x)$.

Now we can apply the Implicit function theorem to $f \circ \varphi$. Write $(u_0, v_0) = \varphi^{-1}(x) \in \mathbb{R}^m \times \mathbb{R}^n$. Then $\frac{\partial(f \circ \varphi)}{\partial v}(u_0, v_0)$ is non-singular. There is an open neighborhood U of u_0 in $\mathbb{R}^m$ and an open neighborhood V of v_0 in $\mathbb{R}^n$ such that

$$(U \times V) \cap (f \circ \varphi)^{-1}(y) = \{(u, h(u)) \mid u \in U\}.$$

The latter set is the graph of the smooth function h on U, therefore it is diffeomorphic to U (see 19.1). On the other hand the former set is $(U \times V) \cap \left(\varphi^{-1}(f^{-1}(y))\right) = \varphi^{-1}\left(\varphi(U \times V) \cap f^{-1}(y)\right)$. Since φ is a diffeomorphism, we can conclude that $\varphi(U \times V) \cap f^{-1}(y)$ is diffeomorphic to U. $\square$

Preimage of a regular value

Let M and N be manifolds and let $f : M \to N$ be smooth. A point in M is called a **critical point** [1] of f if the derivative of f at that point is not surjective. Otherwise the point is called a **regular point** [2] of f.

Example. Let U be an open set in $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be smooth. Then $x \in U$ is a critical point of f if and only if $\nabla f(x) = 0$. This is exactly the notion of critical point in Calculus.

A point in N is called a **critical value** of f if it is the value of f at a critical point. Otherwise the point is called a **regular value** of f. Thus y is a critical value of f if and only if $f^{-1}(y)$ contains a critical point. In particular, if $f^{-1}(y) = \emptyset$ then y is considered a regular value in this convention.

Example. If $f : M \to N$ where $\dim(M) < \dim(N)$ then every $x \in M$ is a critical point and every $y \in f(M)$ is a critical value of f.

With these notions we immediately obtain a consequence of 21.1:

21.2 Proposition. *If y is a regular value of $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ then $f^{-1}(y)$ is a manifold of dimension m.*

On manifolds we have the following results.

21.3 Proposition. *Let M and N be manifolds and let $f : M \to N$ be smooth. If $\dim(M) = \dim(N)$ and y is a regular value of f then $f^{-1}(y)$ is a discrete set, in other words, $f^{-1}(y)$ is a zero dimensional manifold.*

Proof. If $x \in f^{-1}(y)$ then there is a neighborhood of x on which f is a bijection (a consequence of the Inverse function theorem, 21.23). That neighborhood contains no other point in $f^{-1}(y)$. Thus $f^{-1}(y)$ is a discrete set, i.e. each point has a neighborhood containing no other points. □

21.4 Theorem (level set at a regular value is a manifold). *Let M and N be manifolds, $\dim(M) \geq \dim(N)$, and let $f : M \to N$ be smooth. If y is a regular value of $f : M \to N$ then $f^{-1}(y)$ is a manifold of dimension $\dim(M) - \dim(N)$.*

Example. To be able to follow the proof more easily the reader can try to work it out in an example, such as the case where M is the graph of the function $z = x^2 + y^2$, and f is the height function $f((x, y, z)) = z$ defined on M.

[1] điểm dừng, điểm tới hạn

[2] điểm thường, điểm chính quy

Proof. Let $m = \dim(M)$ and $n = \dim(N)$. The case $m = n$ is already considered in 21.3. Now we assume $m > n$. Let $x_0 \in f^{-1}(y_0)$. Consider the diagram

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle\varphi}\uparrow & & \uparrow{\scriptstyle\psi} \\ \mathbb{R}^m \supset O & \xrightarrow[g]{} & W \subset \mathbb{R}^n \end{array}$$

where $g = \psi^{-1} \circ f \circ \varphi$ and $\psi(w_0) = y_0$. Since df_{x_0} is onto, $dg_{\varphi^{-1}(x_0)}$ is also onto. Let $u_0 = \varphi^{-1}(x_0) \in \mathbb{R}^m$. Applying 21.1 to g, there is an open neighborhood U of u_0 in $\mathbb{R}^m$ contained in O and $U \cap g^{-1}(w_0)$ is diffeomorphic to an open set V of $\mathbb{R}^{m-n}$. We can check that

$$\varphi\left(U \cap g^{-1}(w_0)\right) = \varphi(U) \cap f^{-1}(y).$$

Thus $\varphi(U) \cap f^{-1}(y)$, which is an open neighborhood of x_0 in $f^{-1}(y)$, is diffeomorphic to V. $\square$

Example. Consider the subset of $\mathbb{R}^2$ given by the equation

$$x^2 + y^4 = 1.$$

Let $f(x, y) = x^2 + y^4$. Since $\nabla f(x, y) = (2x, 4y^3)$, the only critical point of f is $(0, 0)$, and the only critical value is $f(0, 0) = 0$. Hence 1 is a regular value of f, and the set $f^{-1}(1)$ is a one-dimensional smooth manifold in $\mathbb{R}^2$.

Example. Consider the graph M of a smooth function $f : \mathbb{R} \to \mathbb{R}$. Let $g : \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) = f(x) - y$. Then $M = g^{-1}(0)$. Since $\nabla g(x, y) = (f'(x), -1) \neq 0$, g has no critical points. Thus 0 is a regular value for g, hence $M = g^{-1}(0)$ is a smooth one-dimensional manifold in $\mathbb{R}^2$.

Example. The n-sphere S^n is a subset of $\mathbb{R}^{n+1}$ determined by the implicit equation $\sum_{i=1}^{n+1} x_i^2 = 1$. Since 1 is a regular value of the function $f(x_1, x_2, \ldots, x_{n+1}) = \sum_{i=1}^{n+1} x_i^2$, the sphere S^n is a manifold of dimension n.

Lie groups

Definition. *A* **Lie group** [1] *is a smooth manifold which is also a group, for which the group operations are compatible with the smooth structure, namely the group multiplication and inversion are smooth.*

Here we consider examples of Lie groups coming from linear groups.

The set $M(n)$ of all $n \times n$ matrices whose entries are real numbers can be identified with the Euclidean manifold $\mathbb{R}^{n^2}$, as in 7.31, by considering entries

[1] Sophus Lie was a mathematician living in the 19 century.

of a matrix as coordinates, via a map such as

$$(a_{i,j}) \longmapsto (a_{1,1}, a_{2,1}, \ldots, a_{n,1}, a_{1,2}, a_{2,2}, \ldots, a_{n,2}, a_{1,3}, \ldots, a_{n-1,n}, a_{n,n}).$$

Consider the determinant map $\det : M_n(\mathbb{R}) \to \mathbb{R}$. For $A = [a_{i,j}] \in M(n)$,

$$\det(A) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}.$$

From this we can see that det is a smooth function.

Let us find the critical points of det. Fixing i, we have the formula

$$\det(A) = \sum_j (-1)^{i+j} a_{i,j} \det(A^{i,j}),$$

where $A^{i,j}$ is obtained from A by omitting the ith row and jth column [Lan87, p. 150]. A critical point is a matrix $A = [a_{i,j}]$ such that

$$\frac{\partial \det}{\partial a_{i,j}}(A) = (-1)^{i+j} \det(A^{i,j}) = 0$$

for all i, j. In particular, $\det(A) = 0$. So 0 is the only critical value of det.

Example. The preimage $\mathrm{SL}(n) = \det^{-1}(1)$ is then a manifold of dimension $n^2 - 1$ in $\mathbb{R}^{n^2}$. Under matrix multiplication and inverse, $\mathrm{SL}(n)$ is also a group, called the Special Linear Group. Notice that the group multiplication in $\mathrm{SL}(n)$ is a smooth map from $\mathrm{SL}(n) \times \mathrm{SL}(n)$ to $\mathrm{SL}(n)$. The group inverse operation is a smooth map from $\mathrm{SL}(n)$ to itself, by Formula (19.2). The group $\mathrm{SL}(n)$ is an example of a Lie group.

Next we consider the group $O(n)$ of orthogonal $n \times n$ matrices, that is, $O(n) = \{A \in M(n) \mid A^t A = I\}$, where A^t denotes the transpose matrix of the matrix A.

21.5 Proposition. *The orthogonal group $O(n)$ is the group of all representative matrices of linear transformations of $\mathbb{R}^n$ that preserve distance.*

Proof. The Euclidean inner product can be given by the Euclidean distance

$$\langle x, y \rangle = \frac{1}{2}\left(\|x + y\|^2 - \|x\|^2 - \|y\|^2\right),$$

therefore if A is the matrix of a distance-preserving linear map then

$$\begin{aligned}
\langle Ax, Ay \rangle &= \frac{1}{2}\left(\|Ax + Ay\|^2 - \|Ax\|^2 - \|Ay\|^2\right) \\
&= \frac{1}{2}\left(\|A(x + y)\|^2 - \|Ax\|^2 - \|Ay\|^2\right) \\
&= \frac{1}{2}\left(\|x + y\|^2 - \|x\|^2 - \|y\|^2\right) = \langle x, y \rangle.
\end{aligned}$$

Thus A preserves inner product. Vice versa, if A preserves inner product then A preserves norm, hence $\|Ax - Ay\| = \|A(x-y)\| = \|x-y\|$, thus A preserves distance.

Next we use the identity

$$\langle Ax, Ay \rangle = \langle x, A^t Ay \rangle .$$

Indeed,

$$\begin{aligned}
\langle Ax, Ay \rangle &= \sum_i (Ax)_i (Ay)_i = \sum_i \left(\sum_j A_{i,j} x_j \right) \left(\sum_k A_{i,k} y_k \right) \\
&= \sum_i \sum_k \left(\sum_j A^t_{j,i} A_{i,k} \right) x_j y_k = \sum_j \sum_k \left(\sum_i A^t_{j,i} A_{i,k} \right) x_j y_k \\
&= \sum_j \sum_k (A^t A)_{j,k} x_j y_k = \sum_j x_j \left((A^t A) y \right)_j = \langle x, A^t A y \rangle .
\end{aligned}$$

So A preserves inner product if and only if $\langle x, y \rangle = \langle x, A^t A y \rangle$ for all x, y, which is if and only if $A^t A = I$, that is, A is orthogonal. □

It is known that any isometry of Euclidean spaces is a composition of a translation and an orthogonal transformation, see for instance [Mun91, p. 173], [Vhhvp].

Proposition. *The orthogonal group $O(n)$ is a Lie group.*

Proof. Let $SY(n)$ be the set of symmetric $n \times n$ matrices, which can be identified with the Euclidean space $\mathbb{R}^{\frac{n^2+n}{2}}$. Consider the smooth map $f : M(n) \to SY(n)$, $f(A) = A^t A$. We have $O(n) = f^{-1}(I)$. We will show that I is a regular value of f.

We compute the derivative of f at $A \in f^{-1}(I)$:

$$\begin{aligned}
df_A(B) &= \lim_{h \to 0} \frac{f(A+hB) - f(A)}{h} = \lim_{t \to 0} \frac{(A^t + tB^t)(A+hB) - A^t A}{h} \\
&= \lim_{h \to 0} \frac{hA^t B + hB^t A + h^2 B^t B}{h} = B^t A + A^t B.
\end{aligned}$$

To check whether df_A is onto for $A \in O(n)$, we need to check that given $C \in SY(n)$ there is a $B \in M(n)$ such that $C = B^t A + A^t B$. We can write $C = \frac{1}{2}C + \frac{1}{2}C$, then the equation $\frac{1}{2}C = A^t B$ gives a solution $B = \frac{1}{2}AC$, then $B^t A = \frac{1}{2}C^t A^t A = \frac{1}{2}C^t = \frac{1}{2}C$. So $B = \frac{1}{2}AC$ is indeed a solution to the equation $C = B^t A + A^t B$. Thus I is a regular value of f, and $O(n)$ is a manifold in $\mathbb{R}^{n^2}$, of dimension $\frac{n^2+n}{2} - n = \frac{n^2-n}{2}$, and is a Lie group. □

Problems

21.6. The conic curves could be obtained as intersections between a plane and a cone in $\mathbb{R}^3$. In $\mathbb{R}^2$ they have the following equations

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

$$y^2=2px, p>0$$

Draw the conic curves. Which conic curves are smooth manifolds in $\mathbb{R}^2$?

21.7. Does the equation

$$x^2+5xy+6y^4=25$$

determine a smooth manifold in $\mathbb{R}^2$? Draw it.

21.8. Consider the following quadratic surfaces in $\mathbb{R}^3$,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$$

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$$

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$$

Draw the surfaces. Which of them are smooth manifolds in $\mathbb{R}^3$?

21.9. Let $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=x^2-y^2$. Show that if $a\neq 0$ then $f^{-1}(a)$ is a 1-dimensional manifold, but $f^{-1}(0)$ is not. Show that if a and b are both positive or both negative then $f^{-1}(a)$ and $f^{-1}(b)$ are diffeomorphic.

21.10. Let $f:\mathbb{R}^3\to\mathbb{R}$, $f(x,y,z)=x^2+y^2-z^2$. Show that if $a\neq 0$ then $f^{-1}(a)$ is a 2-dimensional manifold, but $f^{-1}(0)$ is not. Show that if a and b are both positive or both negative then $f^{-1}(a)$ and $f^{-1}(b)$ are diffeomorphic.

21.11. Show that the equation $x^6+y^4+z^2=1$ determine a manifold in $\mathbb{R}^3$.

21.12. Use regular value to show that the torus T^2 is a smooth manifold.

21.13. Find the regular values of the function $f(x,y,z)=[4x^2(1-x^2)-y^2]^2+z^2$. Use a computer program to draw level sets of this function. There is a level for which the level set is a genus two surface.

21.14. Is the intersection of the two surfaces $z=x^2+y^2$ and $z=1-x^2-y$ a manifold?

21.15. Show that the height function $(x,y,z)\mapsto z$ on the sphere S^2 has exactly two critical points.

21.16. Show that a smooth function on a compact manifold must have at least two critical points.

21.17 (derivatives of complex functions). In complex analysis, for a map $f : \mathbb{C} \to \mathbb{C}$ we can consider its complex derivative $f'(z)$. We can also consider f as a map $f : \mathbb{R}^2 \to \mathbb{R}^2$, by writing $z = (x, y) \in \mathbb{R}^2$ and $f(x, y) = (u(x, y), v(x, y) \in \mathbb{R}^2$, then consider its Fréchet derivative $df_{(x,y)}$ which is also its derivative as a map between two smooth manifolds.

(a) Compute $f'(z)$ and $df_{(x,y)}$.

(b) Check that $\det df(x, y) = |f'(z)|^2$.

(c) Are there relations between zeros of f as a complex function and critical points of f as a map between smooth manifolds?

21.18. Show that if M is compact, $\dim(M) = \dim(N)$, and y is a regular value of f, then $f^{-1}(y)$ is a finite set.

21.19. Let $\dim(M) = \dim(N)$, M be compact and S be the set of all regular values of $f : M \to N$. For $y \in S$, let $\left|f^{-1}(y)\right|$ be the number of elements of $f^{-1}(y)$. Show that the map

$$\begin{array}{ccc} S & \to & \mathbb{N} \\ y & \mapsto & \left|f^{-1}(y)\right|. \end{array}$$

is locally constant. In other words, each regular value has a neighborhood where the number of preimages of regular values is constant.

21.20. Let M be a compact manifold and let $f : M \to \mathbb{R}$ be smooth. Show that the set of regular values of f is open.

21.21. Find a counter-example to show that 21.3 is not correct if regular value is replaced by critical value.

21.22. Show that if $f : M \to N$ is smooth, y is a regular of f, and $x \in f^{-1}(y)$, then $\ker df_x = T_x f^{-1}(y)$.

21.23 (Inverse function theorem for manifolds). Let M and N be two manifolds of the same dimensions, and let $f : M \to N$ be smooth. Show that if x is a regular point of f then there is a neighborhood in M of x on which f is a diffeomorphism onto its image.

21.24. Let M be a smooth m-dimensional manifold and suppose that $\varphi : \mathbb{R}^m \to M$ is injective, smooth, and $d\varphi(u)$ has rank m for each $u \in \mathbb{R}^m$. Show that φ is a local parametrization of M (compare 19.4).

For example, once we know that the torus is a smooth manifold, we can deduce that map $\psi(\phi, \theta) = ((b + a\cos\theta)\cos\phi, (b + a\cos\theta)\sin\phi, a\sin\theta)$ with $0 < \theta < 2\pi$ and $0 < \phi < 2\pi$ (see 8.10) is indeed a local parametrization of the torus, just by directly checking that this map is injective and its Jacobian matrix is non-singular. There is no need to write down the inverse map of ψ then check that it is smooth.

21.25. Check that our definition of smooth manifold coincides with the definitions in [Spi65, p. 109] .

21.26. Show that S^1 is a Lie group.

21.27. Show that the set of all invertible $n \times n$-matrices $\mathrm{GL}(n)$ is a Lie group and find its dimension.

21.28. In this problem we find the tangent spaces of $\mathrm{SL}(n)$.

(a) Check that the derivative of the determinant map $\det : M_n(\mathbb{R}) \to \mathbb{R}$ is represented by a gradient vector whose (i, j)-entry is $(-1)^{i+j} \det(A^{i,j})$.

(b) Determine the tangent space of $\mathrm{SL}(n)$ at $A \in \mathrm{SL}(n)$.

(c) Show that the tangent space of $\mathrm{SL}(n)$ at the identity matrix is the set of all $n \times n$ matrices with zero traces.

22 Critical points of real functions

Gradient vector

Let $M \subset \mathbb{R}^k$ be an m-dimensional smooth manifold and let $f : M \to \mathbb{R}$ be smooth.

The tangent space TM_x is a linear subspace of the ambient space $\mathbb{R}^k$, therefore it inherits the Euclidean inner product from $\mathbb{R}^k$. In this inner product space the linear map $df_x : TM_x \to \mathbb{R}$ is represented by an element in TM_x which we called the **gradient vector** $\nabla f(x)$ or $\operatorname{grad} f(x)$, determined uniquely by the property

$$\langle \nabla f(x), v \rangle = df_x(v), \forall v \in TM_x.$$

This is treated in Linear Algebra [Lan87, p. 128], or can be considered as a simple case of the Riesz representation theorem in Functional Analysis [KF75], [TTV]. Notice that the gradient vector field ∇f is defined on the whole manifold, and ***gradient vector does not depending on local coordinates***.

Let U be an open neighborhood in M parametrized by φ. For each $x = \varphi(u) \in U$ we define the **first partial derivative**:

$$\begin{aligned} \left(\frac{\partial}{\partial x_i} f\right) &: U \to \mathbb{R} \\ x = \varphi(u) &\mapsto \left(\frac{\partial}{\partial x_i} f\right)(x) = \frac{\partial}{\partial u_i}(f \circ \varphi)(u). \end{aligned}$$

This definition depends on local coordinates. By this definition, $\left(\frac{\partial}{\partial x_i} f\right) \circ \varphi$ is smooth, therefore $\left(\frac{\partial}{\partial x_i} f\right)$ is smooth (see 19.19).

Example. If $M = \mathbb{R}^m$ and $\varphi = \mathrm{id}$ this is the usual partial derivative.

We can think that the parametrization φ brings the coordinate system of $\mathbb{R}^m$ to the neighborhood U, then $\left(\frac{\partial}{\partial x_i} f\right)(x)$ is the rate of change of f when the variable x changes along the path in U which is the composition of the path te_i along the ith axis of $\mathbb{R}^m$ with φ.

We can write

$$\begin{aligned} \left(\frac{\partial}{\partial x_i} f\right)(x) &= \frac{\partial}{\partial u_i}(f \circ \varphi)(u) = d(f \circ \varphi)(u)(e_i) \\ &= \big(df(x) \circ d\varphi(u)\big)(e_i) = df(x)(d\varphi(u)(e_i)). \end{aligned}$$

Thus $\left(\frac{\partial}{\partial x_i} f\right)(x)$ is the value of the derivative map $df(x)$ at the image of the unit vector e_i of $\mathbb{R}^m$.

In the local parametrization by φ the vectors $d\varphi(u)(e_i) = \frac{\partial \varphi}{\partial u_i}(u)$, $1 \le i \le m$ constitutes a vector basis for TM_x. In this basis the linear map $df(x)$ is represented by a $1 \times m$-matrix whose coordinates are

$$df_x(d\varphi(u)(e_i)) = \frac{\partial f}{\partial x_i}(x).$$

Thus in that basis we have the familiar formula $[df(x)] = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_m}\right)$.

Let us see how ∇f is written in local coordinates. Take the linear basis $\left(\frac{\partial \varphi}{\partial u_1}(u), \dots, \frac{\partial \varphi}{\partial u_m}(u)\right)$ of TM_x. Let $g_{ij} = \left\langle \frac{\partial \varphi}{\partial u_i}, \frac{\partial \varphi}{\partial u_j} \right\rangle$, then $g_{ij} = g_{ji}$ and $g = (g_{ij})_{i,j}$ is a symmetric matrix. In the subject of Differential Geometry the functions g_{ij} are called "coefficients of Riemannian metric". Letting $\nabla f(\varphi(u)) = \sum_{i=1}^m a_i(u) \frac{\partial \varphi}{\partial u_i}(u)$, we have

$$\begin{aligned}\left\langle \nabla f(\varphi(u)), \frac{\partial \varphi}{\partial u_i}(u) \right\rangle &= \sum_{j=1}^m a_j(u) \left\langle \frac{\partial \varphi}{\partial u_j}, \frac{\partial \varphi}{\partial u_i} \right\rangle (u) = \sum_{j=1}^m a_j(u) g_{ji}(u) = \sum_{j=1}^m g_{ij}(u) a_j(u) \\ &= df(x)\left(\frac{\partial \varphi}{\partial u_i}(u)\right) = \frac{\partial f}{\partial x_i}(x).\end{aligned}$$

We get a matrix equation

$$g \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{pmatrix},$$

thus

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = g^{-1} \cdot \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{pmatrix},$$

so in this linear basis the gradient vector is represented by

$$[\nabla f] = g^{-1} \cdot \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{pmatrix}.$$

This formula shows that the gradient $\nabla f : M \to \mathbb{R}^k$ is a smooth function.

As in Calculus, we have the following interpretation of the direction of the gradient vector:

22.1 Proposition. *Let $M \subset \mathbb{R}^k$ be a smooth manifold, let $f : M \to \mathbb{R}$ be smooth, and let c be a regular value of f.*

(a) *Let $N = f^{-1}(c)$, then for any $x \in N$, $\nabla f(x) \perp TN_x$ (in the Euclidean inner product of $\mathbb{R}^k$). Thus **gradient vector is normal to level set**.*

(b) *Among all unit vectors $v \in TM_x$, the number $df_x(v)$ is greatest when $v = \frac{\nabla f(x)}{\|\nabla f(x)\|}$. Thus* ***the direction of the gradient vector is the direction in which the value of the function increases the greatest.***

The opposite direction of the gradient vector is the direction in which the value function decreases the greatest. This is the basis for the method of **gradient descent** in optimization.

Proof. (a) Recall that N is a manifold of dimension $\dim M - 1$. Let $v \in TN_x$, then $v = \alpha'(0)$ is the velocity vector of a smooth path α in N going through $x = \alpha(0)$. Since $f \circ \alpha = c$ is a constant function, the derivative of $f \circ \alpha$ is 0, so $\frac{d}{dt}(f \circ \alpha)(0) = df_x(\alpha'(0)) = df_x(v) = \langle \nabla f(x), v \rangle = 0$, thus $\nabla f(x) \perp v$.

(b) Since $df_x(v) = \langle \nabla f(x), v \rangle$ it is greatest when $v = \frac{\nabla f(x)}{\|\nabla f(x)\|}$. □

Example. Find the tangent space to the surface $\frac{x^2}{4} + \frac{y^2}{9} + \frac{z^2}{16} = 3$ at $(2, 3, 4)$.

Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = \frac{x^2}{4} + \frac{y^2}{9} + \frac{z^2}{16}$. The gradient vector $\nabla f(x, y, z) = (\frac{x}{2}, \frac{2y}{9}, \frac{z}{8})$. The only critical point is $(0, 0, 0)$, so the only critical value is 0. Since 3 is a regular value of f, the preimage $S = f^{-1}(3)$ is a two-dimensional manifold in $\mathbb{R}^3$. The gradient vector $\nabla f(x, y, z) = (\frac{x}{2}, \frac{2y}{9}, \frac{z}{8})$ is normal to S at $(x, y, z) \in S$. Thus the tangent space of S at $(2, 3, 4)$ consists of points $(x, y, z) \in \mathbb{R}^3$ satisfying the equation

$$\left(1, \frac{2}{3}, \frac{1}{2}\right) \cdot (x, y, z) = 0.$$

Meanwhile the tangent plane (in the sense in Calculus) consists of points $(x, y, z) \in \mathbb{R}^3$ satisfying the equation

$$\left(1, \frac{2}{3}, \frac{1}{2}\right) \cdot (x - 2, y - 3, z - 4) = 0,$$

equivalently

$$x + \frac{2}{3}y + \frac{1}{2}z = 6.$$

Critical points and Extrema

On relationships between critical points and local extrema, as in Calculus, we make the following simple observations:

Proposition. *Let $f : M \to \mathbb{R}$ be smooth and let $x \in M$. The point x is a critical point of f if and only if all partial derivatives of f are zero at x, that is $\frac{\partial f}{\partial x_i}(x) = 0, \ \forall i$. Equivalently the point x is a critical point of f if and only if the gradient vector of f is zero at x, that is $\nabla f(x) = 0$.*

Proposition. *A local extremum point is a critical point.*

Proof. Suppose that f has a local extremum at x. There is an open neighborhood U of x in M parametrized by φ with $\varphi(u) = x$. Since $f \circ \varphi$ has a local extremum at u, from Calculus we know u is a critical point of $f \circ \varphi$. Since $d(f \circ \varphi)_u = df_x \circ d\varphi_u$ and $d\varphi_u$ is an isomorphism, we must have $df_x = 0$, thus x is a critical point of f. $\square$

Since $\frac{\partial}{\partial x_i} f$ is a smooth function on U, we can take its partial derivatives. Thus we define the **second partial derivatives**:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f\right)(x).$$

In other words,

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i \partial x_j}(\varphi(u)) &= \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f\right)(\varphi(u)) = \frac{\partial}{\partial u_i}\left(\frac{\partial f}{\partial x_j} \circ \varphi\right)(u) \\
&= \frac{\partial}{\partial u_i}\left(\frac{\partial}{\partial u_j}(f \circ \varphi)\right)(u) = \frac{\partial^2}{\partial u_i \partial u_j}(f \circ \varphi)(u).
\end{aligned}$$

Consider the Hessian matrix of second partial derivatives:

$$Hf(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{1 \le i,j \le m}.$$

If the Hessian matrix of f at x is non-singular then we say that x is a **non-degenerate critical point** of f.

Lemma. *The non-degeneracy of a critical point does not depend on choices of local coordinates.*

Proof. The problem is reduced to the case of functions on $\mathbb{R}^m$. If $f : \mathbb{R}^m \to \mathbb{R}$ and φ is a change of variables i.e. a diffeomorphism of $\mathbb{R}^m$ then

$$\frac{\partial}{\partial u_i}(f \circ \varphi)(u) = \sum_k \frac{\partial f}{\partial x_k}(x) \cdot \frac{\partial \varphi_k}{\partial u_i}(u).$$

From this we observe that if x is a critical point of f then u is a critical point of $f \circ \varphi$. Next, at the critical point x of f:

$$\begin{aligned}
\frac{\partial^2}{\partial u_j \partial u_i}(f \circ \varphi)(u) &= \sum_k \left[\left(\sum_l \frac{\partial^2 f}{\partial x_l \partial x_k}(x) \cdot \frac{\partial \varphi_l}{\partial u_j}(u)\right) \cdot \frac{\partial \varphi_k}{\partial u_i}(u) + \frac{\partial f}{\partial x_k}(x) \cdot \frac{\partial^2 \varphi_k}{\partial u_j \partial u_i}(u)\right] \\
&= \sum_{k,l} \frac{\partial^2 f}{\partial x_l \partial x_k}(x) \cdot \frac{\partial \varphi_l}{\partial u_j}(u) \cdot \frac{\partial \varphi_k}{\partial u_i}(u).
\end{aligned}$$

Rewriting in matrix notation:

$$H(f \circ \varphi)(u) = J\varphi(u)^t \cdot Hf(\varphi(u)) \cdot J\varphi(u). \tag{22.2}$$

Taking determinant of both sides we see that $H(f \circ \varphi)(u)$ is degenerate if and only if $Hf(\varphi(u))$ is degenerate. □

Morse lemma

Relating to the second derivative test for extrema in Calculus, we have the following result.

Theorem (Morse lemma [1]). *Suppose that $f : M \to \mathbb{R}$ is smooth and p is a non-degenerate critical point of f. There is a local coordinate φ in a neighborhood of p such that in that neighborhood*

$$(f \circ \varphi)(u) = f(p) - u_1^2 - u_2^2 - \cdots - u_k^2 + u_{k+1}^2 + u_{k+2}^2 + \cdots + u_m^2.$$

The number k does not depend on the choice of such local coordinate and is called the ***index of the non-degenerate critical point*** *p.*

Proof. Since we only need to prove the formula for $f \circ \varphi^{-1}$, we only need to work in $\mathbb{R}^m$.

First we show that we can write

$$f(x) = f(0) + \sum_{i,j=1}^{m} x_i x_j h_{i,j}(x)$$

where $h_{i,j} = h_{j,i}$. This looks very much like a Taylor expansion, however here we want to make sure $h_{i,j}$ are smooth.

Let us write

$$\begin{aligned} f(x) &= f(0) + \int_0^1 \frac{d}{dt} f(tx)\, dt \\ &= f(0) + \sum_{i=1}^{m} \int_0^1 \left(\frac{\partial f}{\partial x_i}(tx) \right) x_i \, dt \\ &= f(0) + \sum_{i=1}^{m} x_i \int_0^1 \left(\frac{\partial f}{\partial x_i}(tx) \right) dt \\ &= f(0) + \sum_{i=1}^{m} x_i g_i(x), \end{aligned}$$

where the functions $g_i(x) = \int_0^1 \left(\frac{\partial}{\partial x_i} f(tx) \right) dt$ are smooth (which we can check,

[1] Marston Morse studied critical points of smooth functions on manifolds from the 1920s.

or see for example [Lan97, p. 276]). Notice that $g_i(0) = \frac{\partial f}{\partial x_i}(0) = 0$. Apply the above construction once again to g_i we obtain smooth functions $g_{i,j}$ such that

$$f(x) = f(0) + \sum_{i,j=1}^{m} x_i x_j g_{i,j}(x).$$

Set $h_{i,j} = (g_{i,j} + g_{j,i})/2$ then $h_{i,j} = h_{j,i}$ and $f(x) = f(0) + \sum_{i,j=1}^{m} x_i x_j h_{i,j}(x)$.

The rest of the proof is completing the square. Notice that $h_{i,j}(0) = \frac{1}{2}\frac{\partial^2 f}{\partial x_i \partial x_j}(0)$. First we show that we can assume that $h_{1,1}(0) \neq 0$. Indeed, suppose that $h_{1,1}(0) = 0$. Since the matrix $\left(h_{i,j}(0)\right)$ is non-singular, there is an index $i_0 \neq 1$ such that $h_{1,i_0}(0) \neq 0$. Make the change of variable $x_{i_0} = x'_{i_0} + x_1$, then in this new variable $h'_{1,1}(0) = 2h_{1,i_0}(0) \neq 0$.

When $h_{1,1}(0) \neq 0$ there is a neighborhood of 0 such that $h_{1,1}(x)$ does not change its sign. In that neighborhood, if $h_{1,1}(0) > 0$ then

$$\begin{aligned}
f(x) &= f(0) + \quad h_{1,1}(x)x_1^2 + \sum_{1<j}(h_{1,j}(x) + h_{j,1}(x))x_1 x_j + \sum_{1<i,j} h_{i,j}(x) x_i x_j \\
&= f(0) + \quad h_{1,1}(x)x_1^2 + 2\sum_{1<j} h_{1,j}(x) x_1 x_j + \sum_{1<i,j} h_{i,j}(x) x_i x_j \\
&= f(0) + \quad \left(\sqrt{h_{1,1}(x)}x_1\right)^2 + 2\sqrt{h_{1,1}(x)}x_1 \frac{\sum_{1<j} h_{1,j}(x) x_j}{\sqrt{h_{1,1}(x)}} + \sum_{1<i,j} h_{i,j}(x) x_i x_j \\
&= f(0) + \quad \left[\sqrt{h_{1,1}(x)}x_1 + \sum_{1<j} \frac{h_{1,j}(x)}{\sqrt{h_{1,1}(x)}} x_j\right]^2 \\
&\qquad - \left(\sum_{1<j} \frac{h_{1,j}(x)}{\sqrt{h_{1,1}(x)}} x_j\right)^2 + \sum_{1<i,j} h_{i,j}(x) x_i x_j.
\end{aligned}$$

Similarly, if $h_{1,1}(0) < 0$ then

$$\begin{aligned}
f(x) &= f(0) - \quad \left[\sqrt{-h_{1,1}(x)}x_1 - \sum_{1<j} \frac{h_{1,j}(x)}{\sqrt{-h_{1,1}(x)}} x_j\right]^2 \\
&\qquad + \left(\sum_{1<j} \frac{h_{1,j}(x)}{\sqrt{-h_{1,1}(x)}} x_j\right)^2 + \sum_{1<i,j} h_{i,j}(x) x_i x_j.
\end{aligned}$$

Combining both cases, we define the new variables:

$$\begin{aligned}
v_1 &= \sqrt{|h_{1,1}(x)|}x_1 + \operatorname{sign}(h_{1,1}(0)) \sum_{1<j} \frac{h_{1,j}(x)}{\sqrt{|h_{1,1}(x)|}} x_j, \\
v_i &= x_i, \ i > 1.
\end{aligned}$$

Computing derivative directly, we get

$$\frac{\partial v_1}{\partial x_1}(0) = \sqrt{|h_{1,1}(0)|} \neq 0$$

so the Jacobian matrix $\left(\frac{\partial v_i}{\partial x_j}(0)\right)$ is non-singular. By the Inverse function theorem, there is a neighborhood of 0 where the correspondence $x \mapsto v$ is a diffeomorphism, that is, a change of variables. With the new variables we have

$$f(v) = f(0) + \operatorname{sign}(h_{1,1}(0))v_1^2 + \sum_{1<i,j} h'_{i,j}(v)v_iv_j.$$

By a direct calculation, we can check that in these variables

$$Hf(0) = \begin{pmatrix} \operatorname{sign}(h_{1,1}(0)) & 0 \\ 0 & (2h'_{i,j}(0))_{1<i,j\leq m} \end{pmatrix}.$$

Using (22.2) we conclude that the matrix $(h'_{i,j}(0))_{1<i,j\leq m}$ must be non-singular. Thus the induction process can be carried out. Finally we can permute the variables such that in the final form of f the negative signs are in front.

We now show that the number k of negative signs depend only on the function f and does not depend on the choice of change of variable. Suppose that φ is a change of variables with $\varphi(0) = 0$. The number k is precisely the number of negative eigenvalues of $H(f \circ \varphi)(0)$. The Sylvester law on quadratic forms in Algebra (see e.g. [Lan97, p. 577]) says that the numbers of negative eigenvalues of $H(f\circ\varphi)(0)$ and $Hf(0)$ are equal, as the corresponding quadratic forms are related by a change of variable.

For convenience here we provide a sketch of a proof for the Sylvester law. Let V be a subspace of $\mathbb{R}^m$ of the maximum dimension on which the bilinear form represented by the matrix $H(f \circ \varphi)(0)$ is negatively definite, i.e. $v^t \cdot H(f \circ \varphi)(0) \cdot v < 0$ for all $0 \neq v \in V$. Recall (22.2), we have

$$H(f \circ \varphi)(0) = J\varphi(0)^t \cdot Hf(\varphi(0)) \cdot J\varphi(0) = J\varphi(0)^t \cdot Hf(0) \cdot J\varphi(0).$$

Since

$$v^t \cdot H(f\circ\varphi)(0)\cdot v = v^t \cdot J\varphi(0)^t \cdot Hf(0)\cdot J\varphi(0)\cdot v = (J\varphi(0)\cdot v)^t \cdot Hf(0)\cdot (J\varphi(0)\cdot v),$$

we deduce that $Hf(0)$ is negatively definite on $d\varphi(0)(V)$. Thus the maximum dimension of a negatively definite space for $Hf(0)$ is bigger than or equal to the maximum dimension of a negatively definite space for $H(f \circ \varphi)(0)$. By symmetry (namely, by composing with φ^{-1} and using the same argument) the two maximum dimensions must be same, equal to a number l. Since $H(f \circ \varphi)(0)$ is a diagonal matrix with k negative items it is easy to see that

$l \geq k$. A similar argument gives that the maximum dimension of a positively definite space for $H(f \circ \varphi)(0)$ is larger than or equal to $(m-k)$. This implies that $l = k$. □

Example. Non-degenerate critical points of indexes 0 are local minima, ones of indexes m are local maxima, and ones of indexes larger than 0 and less than m are neither minima nor maxima.

Example. On the the sphere $x^2 + y^2 + z^2 = 1$ consider the height function $f(x,y,z) = z$. The upper-half of the sphere is the graph of the function $z = \sqrt{1-x^2-y^2}$. With the local parametrization $\varphi(x,y) = (x, y, \sqrt{1-x^2-y^2})$, we get $f \circ \varphi(x,y) = \sqrt{1-x^2-y^2}$, $\frac{\partial(f\circ\varphi)}{\partial x}(x,y) = -x(1-x^2-y^2)^{-1/2}$, and $\frac{\partial(f\circ\varphi)}{\partial y}(x,y) = -y(1-x^2-y^2)^{-1/2}$. The only critical point of f in the upper-half sphere is $(0,0,1)$. With this local parametrization we calculate the second derivatives,

$$\frac{\partial^2 f}{\partial x^2}(x,y,z) = \frac{\partial^2(f\circ\varphi)}{\partial x^2}(x,y) = -\frac{1-y^2}{(1-x^2-y^2)^{\frac{3}{2}}}$$

$$\frac{\partial^2 f}{\partial x \partial y}(x,y,z) = \frac{\partial^2(f\circ\varphi)}{\partial x \partial y}(x,y) = -\frac{xy}{(1-x^2-y^2)^{\frac{3}{2}}}$$

$$\frac{\partial^2 f}{\partial y^2}(x,y,z) = \frac{\partial^2(f\circ\varphi)}{\partial y^2}(x,y) = -\frac{1-x^2}{(1-x^2-y^2)^{\frac{3}{2}}}.$$

We find

$$Hf(0,0,1) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

So $(0,0,1)$ is a non-degenerate critical point. The index of this critical point is the number of negative eigenvalues of the Hessian matrix, which is 2.

On the lower-half of the sphere we just take the opposite sign for f, we find that f has a non-degenerate critical point on the at $(0,0,-1)$, of index 0.

As an application, we deduce the following sufficient condition for local extrema in Calculus:

Proposition. *Suppose that $f : M \to \mathbb{R}$ is smooth where M is a smooth manifold and x is a critical point of f. The complex eigenvalues of the matrix $Hf(x)$ are non-zero real numbers.*

(a) If all eigenvalues $Hf(x)$ are positive then f has a local minimum at x.

(b) If all eigenvalues $Hf(x)$ are negative then f has a local maximum at x.

(c) If $Hf(x)$ has both positive and negative eigenvalues then f does not have a local extremum at x.

Problems

22.3. Let $f : M \to \mathbb{R}$ be smooth where M is a compact smooth manifold of dimension $m \geq 1$. Show that f has at least two critical points.

22.4. Let $f : M \to \mathbb{R}$ be smooth. Show that $\nabla f = 0$ if and only if f is constant on each connected component of M.

22.5. Show that the following equation determines a surface (i.e. a two dimensional smooth manifold) in $\mathbb{R}^3$:

$$x^2 + y^2 + z^2 + x^4y^4 + x^4z^4 + y^4z^4 - 9z = 21$$

Give an equation for the tangent plane of this surface at the point $(1, 1, 2)$.

22.6. On the the sphere $x^2 + y^2 + z^2 = 1$ consider the height function $f(x, y, z) = z$. Find ∇f.

22.7 (method of Lagrange multipliers). Suppose that $f : M \to \mathbb{R}$ is smooth. Suppose that $g : M \to \mathbb{R}$ is smooth, c a regular value of g, and $N = g^{-1}(c)$. Suppose that $f|_N$ has a local extremum at $x \in N$.

(a) Show that $\nabla f(x) \perp TN_x$ and $\nabla g(x) \perp TN_x$

(b) Deduce that there is $\lambda \in \mathbb{R}$ such that $\nabla f(x) = \lambda \nabla g(x)$.

In Calculus, the above statement can be phrased as the method of Lagrange multipliers: To find the maximum and minimum values of f subject to the constraint $g(x) = c$ (assuming that these extreme values exist and $\nabla g(x) \neq 0$ on the set $g(x) = c$) we need to solve the system $\nabla f(x) = \lambda \nabla g(x)$ and $g(x) = c$ for x and λ.

22.8. * Generalize the method of Lagrange multipliers to two constraints.

22.9. Prove the Morse lemma for the special case of a quadratic function $f(x) = \sum_{1 \leq i,j \leq m} a_{i,j} x_i x_j$, where $a_{i,j}$ are real numbers. In this case we can use a diagonalization of a symmetric matrix or a quadratic form considered in Linear Algebra. The change of variable corresponds to using a new vector basis consisting of eigenvectors of the matrix.

22.10. Show that the set of non-degenerate critical points of a smooth function is discrete, i.e. each element of the set has a neighborhood in the manifold containing no other element of the set (explaining the name, the subspace topology which this set inherits from the manifold coincide with the discrete topology).

22.11 (the classification of critical points in Calculus). Recover the following sufficient condition for extrema in Calculus from the Morse lemma. Suppose that $f : U \to \mathbb{R}$ is smooth where U is an open set in $\mathbb{R}^2$ and (a, b) is a critical point of f. Compute

$$\det Hf(a, b) = f_{xx}(a, b) f_{yy}(a, b) - \left(f_{xy}(a, b)\right)^2.$$

(a) If $\det Hf(a, b) > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a local minimum.

(b) If $\det Hf(a,b) > 0$ and $f_{xx}(a,b) < 0$, then $f(a,b)$ is a local maximum.

(c) If $\det Hf(a,b) < 0$, then $f(a,b)$ is not a local maximum or minimum.

22.12. Suppose that $f : M \to \mathbb{R}$ is smooth where M is a smooth manifold and x is a critical point of f. Let $B(u,v) = u^t \cdot Hf(x) \cdot v$ for $u, v \in \mathbb{R}^m$. Then B is a quadratic form on $\mathbb{R}^m$. Prove that

(a) If B is a positive definite, i.e. $B(u,u) > 0$ for all $u \neq 0$ then f has a local minimum at x.

(b) If B is a negative definite, i.e. $B(u,u) < 0$ for all $u \neq 0$ then f has a local maximum at x.

(c) If there are $u \neq 0$ such that $B(u,u) > 0$ and $v \neq 0$ such that $B(v,v) < 0$ then f does not have a local extremum at x.

22.13 (height function). Let $M \subset \mathbb{R}^k$ be a smooth manifold. Let $v \in \mathbb{R}^k$ with unit length. Consider the function f on M given by $f(x) = \langle x, v\rangle$ (the projection to v) where the inner product is that of $\mathbb{R}^k$. This is the height function in the direction of v. Show that x is a critical point of f if and only if $v \perp TM_x$.

22.14. Consider the torus in $\mathbb{R}^3$ given by the equation $\left(\sqrt{x^2+y^2} - b\right)^2 + z^2 = a^2$, $0 < a < b$. Let $f(x,y,z) = x$ (height function in the direction of the x-axis). Find the non-degenerate critical points of f and their indexes.

23 Flows

Ordinary differential equations

Let V be a smooth vector field defined on an open subset $U \subset \mathbb{R}^m$, i.e. a smooth map $V : U \to \mathbb{R}^m$. An ***integral curve*** of V at a point $x_0 \in U$ is a smooth path $\gamma : (a, b) \to U$ such that $0 \in (a, b)$, $\gamma(0) = x_0$, and $\gamma'(t) = V(\gamma(t))$ for all $t \in (a, b)$. It is a path going through x and at every moment takes the vector of the given vector field as its velocity vector. An integral curve of a vector field is tangent to the vector field. We can consider an integral curve as a solution to the differential equation $\gamma'(t) = V(\gamma(t))$ in $\mathbb{R}^m$ subjected to the initial condition $\gamma(0) = x_0$. In more common notations in differential equations, a local integral curve is a local solution to the following initial value problem:

$$\begin{cases} \frac{dx}{dt} = V(x), \\ x(0) = x_0. \end{cases} \tag{23.1}$$

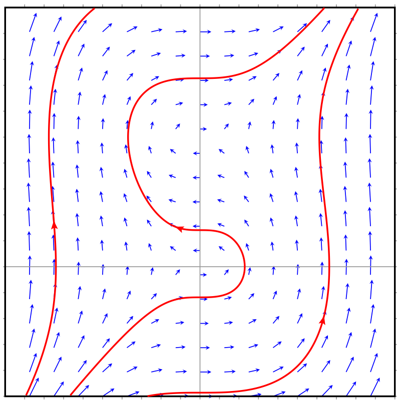


Figure 23.2: A flow on the plane associated with a vector field, with several integral curves.

The following theorems are fundamental results in the subject of Ordinary Differential Equations (ODE) on the existence, uniqueness, and dependence on initial conditions of solutions to differential equations.

Theorem (existence and uniqueness). *If V is smooth in a neighborhood of x_0 then the equation (23.1) has a unique local solution, that is, there is a smooth function x defined on an interval I_x that satisfies (23.1), and if y is a solution defined on an interval I_y then $x = y$ on $I_x \cap I_y$.*

Sketch of Proof. The differential equation (23.1) is equivalent to the integral equation

$$x(t) = x_0 + \int_0^t V(x(s))\, ds. \tag{23.3}$$

Here the integral applies to each component of the vector-valued function. Let T be the map on the right side of (23.3), so T is the map bringing the function x to the function $T(x)$ defined by

$$T(x)(t) = x_0 + \int_0^t V(x(s))\, ds.$$

The main idea of the proof is that (23.3) is equivalent to that the map T has a fixed point, which is then obtained from the contraction mapping theorem (also known as the Banach fixed point theorem) after taking a sufficiently small interval of t so that T becomes a contraction map. We carry out this idea below.

The smooth function V is locally Lipschitz, meaning that there is an open set $W \subset \mathbb{R}^m$ containing x_0 and a real number L such that $\forall a, b \in W, \|V(a) - V(b)\| \leq L\|a-b\|$ (this can be seen as an application of the mean value theorem). There is a real number $r > 0$ such that $B'(x_0, r) \subset W$. Since V is continuous on $B'(x_0, r)$ it is bounded there by a real number $M > 0$. If $x : [-\alpha, \alpha] \to B'(x_0, r)$ then

$$\begin{aligned}\|T(x)(t) - x_0\| &= \| \int_0^t V(x(s))\, ds\| \leq \sqrt{m} \int_0^t \|V(x(s))\|\, ds \\ &\leq \sqrt{m} \int_0^\alpha \|V(x(s))\|\, ds \leq \sqrt{m}\alpha M,\end{aligned}$$

so $T(x)(t) \in B'(x_0, r)$ if $\alpha > 0$ is sufficiently small.

Consider the metric space $C([-\alpha, \alpha], B'(x_0, r))$ of continuous functions $x : [-\alpha, \alpha] \to B'(x_0, r)$ as a closed subspace of $C([-\alpha, \alpha], \mathbb{R}^m)$ under the norm $\|x\| = \sup_{t\in[-\alpha,\alpha]} \|x(t)\|$, which is known to be complete from Functional Analysis. In this space we have the estimates:

$$\begin{aligned}\|T(x) - T(y)\| &= \sup_{t\in[-\alpha,\alpha]} \|(T(x) - T(y))(t)\| = \| \int_0^t [V(x(s)) - V(y(s))]\, ds\| \\ &\leq \sqrt{m} \int_0^\alpha \|V(x(s)) - V(y(s))\|\, ds \leq \sqrt{m} \int_0^\alpha L\|x(s) - y(s)\|\, ds \\ &\leq \sqrt{m}\alpha L||x - y||.\end{aligned}$$

Thus if we take α sufficiently small then T becomes a contraction map on the complete metric space $C([-\alpha, \alpha], B'(x_0, r))$, therefore has a unique solution. □

Let $I = \bigcup_{x\in S} I_x$ where the union is over the set S of all solutions to (23.1). Let $t \in I$, then there is $x \in S$ such that $t \in I_x$. Define $\tilde{x} : I \to U$ by $\tilde{x}(t) = x(t)$. Then $\tilde{x}(t)$ is defined not depending on the choice of x by uniqueness of solution. Furthermore $\tilde{x}$ is also a solution itself, called the maximal solution of (23.1), as it has the maximal domain.

Given a smooth vector field V, for each $x \in U$, let $\phi(t,x)$ or $\phi_t(x)$ be the maximal integral curve of V at x, with t belonging to an interval I_x. We have a map

$$\begin{aligned}\phi : D = \{(t,x) \mid x \in U,\ t \in I_x\} \subset \mathbb{R} \times U &\to U\\ (t,x) &\mapsto \phi_t(x),\end{aligned}$$

with the properties $\phi_0(x) = x$, and $\frac{d}{dt}\phi_t(x) = V(\phi_t(x))$. This map ϕ is called the *flow* (dòng) generated by the vector field V. We can think of ϕ_t as moving every point along the integral curve for an amount of time t.

Theorem (dependence on initial condition). *The flow of a smooth vector field has an open domain and is smooth.*

For proofs and more details see [R. Nagle, E. Saff, A. Snider, *Fundamentals of differential equations and boundary value problems*, 7th ed., tr. 794], [HS74, p. 163, p. 175, p. 302], [Lan97, p. 542, p. 560], [L. Perko, *Differential Equations and Dynamical Systems*, 2001, Chapter 2].

Flows on manifolds

Definition. *A smooth tangent* **vector field** *on a manifold $M \subset \mathbb{R}^k$ is a smooth map $V : M \to \mathbb{R}^k$ such that $V(x) \in TM_x$ for each $x \in M$.*

Example. If $f : M \to \mathbb{R}$ is smooth then the gradient ∇f is a smooth vector field on M.

In a local parametrized neighborhood around a point on the manifold, a smooth vector field on that neighborhood corresponds to a smooth vector field on an open subset of $\mathbb{R}^m$, and an integral curve in that neighborhood corresponds to an integral curve on $\mathbb{R}^m$. Namely, let $\varphi : U \subset \mathbb{R}^m \to M$ be a local parametrization at $x_0 = \varphi(u_0)$ then the corresponding ODE in $\mathbb{R}^m$ is:

$$\begin{cases}\frac{du}{dt} = \left(d\varphi_u\right)^{-1}\left(V(\varphi(u))\right),\\ u(0) = u_0.\end{cases} \tag{23.4}$$

This system has a unique maximal solution u defined on an interval I_u. We check that $\varphi \circ u$ is a solution to 23.1. First,

$$\varphi \circ u(0) = \varphi(u(0)) = \varphi(u_0) = x_0.$$

Next,

$$\frac{d}{dt}(\varphi \circ u)(t) = d\varphi_{u(t)}(u'(t)) = d\varphi_{u(t)}\left(d\varphi_{u(t)}\right)^{-1}\left(V(\varphi(u(t)))\right)$$
$$= V(\varphi(u(t))) = V(\varphi \circ u(t)),$$

as expected. We have just shown that 23.1 has a local solution.

Next we check that this solution does not depend on the choice of local parametrization. Suppose that ψ is another local parametrization and with it a solution $\psi \circ v$ is found. We show that $(\varphi^{-1} \circ \psi) \circ v$ is a solution to 23.4. First, $(\varphi^{-1} \circ \psi) \circ v(0) = (\varphi^{-1} \circ \psi)(v_0) = \varphi^{-1}(\psi(v_0)) = \varphi^{-1}(x_0) = u_0$. Next, by the chain rule

$$\left[(\varphi^{-1} \circ \psi) \circ v\right]'(t) = d(\varphi^{-1} \circ \psi)_{v(t)}(v'(t)) = d(\varphi^{-1} \circ \psi)_{v(t)}\left[\left(d\psi_v\right)^{-1}\left(V(\psi(v(t))\right)\right]$$
$$= d\varphi^{-1}_{\psi(v(t))}\left(V(\psi(v(t))\right) = \left(d\varphi_{(\varphi^{-1}\circ\psi)\circ v(t)}\right)^{-1}\left(V(\psi(v(t))\right)$$
$$= \left(d\varphi_{(\varphi^{-1}\circ\psi)\circ v(t)}\right)^{-1}\left(V(\varphi((\varphi^{-1} \circ \psi) \circ v(t))\right),$$

as expected. As u is already the maximal solution, the solution $(\varphi^{-1} \circ \psi) \circ v$ must agree with u on its domain. Therefore $\psi \circ v$ must agree with $\varphi \circ u$ on its domain. We deduce that $\varphi \circ u$ is the maximal solution for 23.1.

Now for each $x \in M$ let $\phi(t, x)$ or $\phi_t(x)$ be the maximal integral curve of V at x, with t belonging to an interval I_x. We have a map

$$\begin{aligned} \phi : D = \{(t, x) \mid x \in M,\ t \in I_x\} \subset \mathbb{R} \times M &\to M \\ (t, x) &\mapsto \phi_t(x), \end{aligned}$$

with the properties $\phi_0(x) = x$, and $\frac{d}{dt}\phi(t, x) = V(\phi(t, x))$. This map ϕ is called the ***flow*** (dòng) generated by the vector field V. Since a flow on manifold is locally a flow in Euclidean space, we have:

Proposition. *The flow of a smooth vector field on a smooth manifold has an open domain and is smooth.*

Proposition (group law). *Given $x \in M$, if $\phi_t(x)$ is defined on $(-\epsilon, \epsilon)$ then for $s \in (-\epsilon, \epsilon)$ such that $s + t \in (-\epsilon, \epsilon)$ we have*

$$\phi_{t+s}(x) = \phi_t(\phi_s(x)).$$

Proof. Define $\gamma(t) = \phi_{t+s}(x)$. Then $\gamma(0) = \phi_s(x)$, and $\gamma'(t) = \frac{d}{dt}\phi(t + s, x) = V(\phi(t + s, x)) = V(\gamma(t))$. Thus γ is an integral curve at $\phi_s(x)$. But $\beta(t) = \phi_t(\phi_s(x))$ is another integral curve at $\phi_s(x)$. By uniqueness of integral curves, γ must agree with β on their common domains. ∎

Proposition. *If $\phi_t : M \to M$ is defined for $t \in (-\epsilon, \epsilon)$ then it is a diffeomorphism.*

Proof. Since the flow ϕ is smooth the map ϕ_t is smooth. Its inverse map ϕ_{-t} is also smooth. □

When every integral curve can be extended without bound in both directions, that is, for every $x \in M$ the point $\phi_t(x)$ is defined for every $t \in \mathbb{R}$, we say that the flow is **complete**.

Theorem. *On a compact manifold any flow is complete.*

Proof. Although at first each integral curve has its own domain, we will show that for a compact manifold all integral curves can have same domains. Since the domain D of the flow can be taken to be an open subset of $\mathbb{R} \times M$, each $x \in M$ has an open neighborhood U_x and a corresponding interval $(-\epsilon_x, \epsilon_x)$ such that $(-\epsilon_x, \epsilon_x) \times U_x$ is contained in D. The collection $\{U_x \mid x \in M\}$ is an open cover of M, so there is a finite subcover corresponding to x_i, $1 \le i \le n$. Take $\epsilon = \min\{\epsilon_i \mid 1 \le i \le n\}$, then for every $x \in M$ the integral curve $\phi_t(x)$ is defined on $(-\epsilon, \epsilon)$.

Now $\phi_t(x)$ can be extended inductively by intervals of length $\epsilon/2$ to be defined on $\mathbb{R}$. For example, if $t > 0$ then there is $n \in \mathbb{N}$ such that $n\frac{\epsilon}{2} \le t < (n+1)\frac{\epsilon}{2}$, and if $n \ge 2$ we define

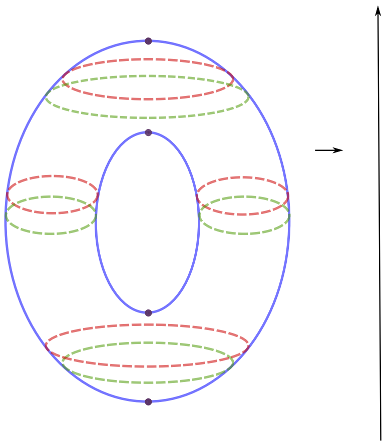
$$\phi_t(x) = \phi_{t-n\frac{\epsilon}{2}}\left(\phi_{n\frac{\epsilon}{2}}(x)\right),$$

where inductively $\phi_{n\frac{\epsilon}{2}}(x) = \phi_{\frac{\epsilon}{2}}\left(\phi_{(n-1)\frac{\epsilon}{2}}(x)\right)$. □

23.5 Theorem. *Let M be a compact smooth manifold and let $f : M \longrightarrow \mathbb{R}$ be smooth. If the interval $[a, b]$ only contains regular values of f then the level sets $f^{-1}(a)$ and $f^{-1}(b)$ are diffeomorphic.*

Briefly, if between the two levels there is no critical point then the two levels are diffeomorphic.

Example. On the torus $(\sqrt{x^2+y^2}-2)^2 + z^2 = 1$, consider the height function along one of its symmetric axes, $f(x, y, z) = x$. From 22.14 we find the critical points are at $y = z = 0$ and $x = -3, -1, 1, 3$, and the critical values are $-3, -1, 1, 3$. For values between -1 and 1 each level set consists of two circles, whereas for values between 1 and 3 each level set consist of one circle.



Proof. The idea is to construct a diffeomorphism from $f^{-1}(a)$ to $f^{-1}(b)$ by pushing points from lower level to higher level along the flow lines of the gradient vector field, as we know that the gradient vector points the direction in which the level increases the greatest (see 22.1). To do this we want the gradient vector at each point to be non-zero. Also we need a complete flow in order to follow the flow lines as long as necessary. We want all the flow lines from one level to reach the other level at the same time, so we want the speed of change of the level $f(\phi_t(x))$ with respect to time to be constant for all x.

By 21.20 there is an interval $(h,k) \supset [a,b]$ such that (h,k) contains only regular values of f. On the open submanifold $N = f^{-1}((h,k))$ the gradient vector ∇f never vanish. By 23.11 there is a smooth function ψ that is 1 on the compact set $f^{-1}([a,b])$ and is 0 outside of N. Let $F = \psi \frac{\nabla f}{||\nabla f||^2}$, then F is a well-defined smooth vector field on M, basically a re-scale of ∇f.

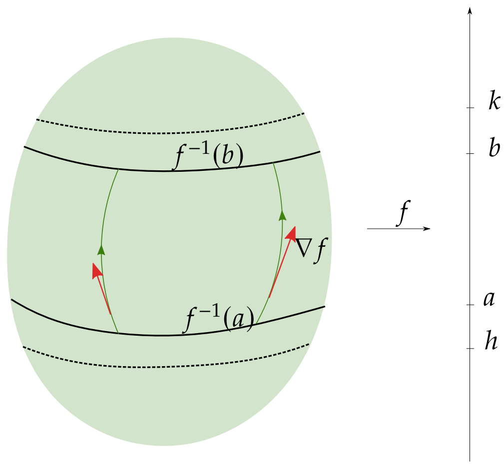


Figure 23.6: The gradient flow going from $f^{-1}(a)$ to $f^{-1}(b)$.

Let ϕ be the flow generated by F. Since M is compact the flow ϕ is complete.

Fix $x \in f^{-1}(a)$, on level a. We calculate, for $t \in \mathbb{R}$,

$$\begin{aligned}\frac{d}{dt}f\left(\phi_t(x)\right) &= df_{\phi_t(x)}\left(\frac{d}{dt}\phi_t(x)\right) = \left\langle \nabla f(\phi_t(x)), F(\phi_t(x))\right\rangle \\ &= \left\langle \nabla f(\phi_t(x)), \psi(\phi_t(x))\frac{\nabla f(\phi_t(x))}{\left\|\nabla f(\phi_t(x))\right\|^2}\right\rangle = \psi(\phi_t(x)).\end{aligned}$$

Thus $\frac{d}{dt}f\left(\phi_t(x)\right) \le 1$ for all $t \in \mathbb{R}$, so $f(\phi_t(x)) \le t + a$. This implies for $t \in [0, b-a]$ the point $\phi_t(x)$ remains in $f^{-1}([a,b])$, thus $\frac{d}{dt}f\left(\phi_t(x)\right) = 1$, thus $f(\phi_t(x)) = t + a$, so $f(\phi_{b-a}(x)) = b$. Hence ϕ_{b-a} maps $f^{-1}(a)$ to $f^{-1}(b)$, is the desired diffeomorphism. □

This theorem together with the Morse lemma are starting results for a theory studying relationships between critical points of real functions and topology of manifolds, called **Morse theory**, see [J. Milnor, *Morse theory*, 1963] and [Y. Matsumoto, *An introduction to Morse theory*, 2002].

23.7 Theorem. *On a connected manifold there is a self diffeomorphism that brings any given point to any given point.*

We say that any connected manifold is ***homogeneous*** (đồng nhất).

Proof. The following proof follows [Mil97]. First we show that we can locally bring any point to any given point without outside disturbance. That translates to a problem on $\mathbb{R}^n$: we will show that for any $c \in B(0,1)$ there is a diffeomorphism $h : \mathbb{R}^n \to \mathbb{R}^n$ such that $h|_{\mathbb{R}^n \setminus B(0,1)} = \mathrm{id}$ and $h(0) = c$.

Take $\|c\| < \delta < 1$. By 23.9 there is a smooth function $\varphi : \mathbb{R}^n \to [0,1]$ such that $\varphi|_{B'(0,\delta)} = 1$ and $\varphi|_{\mathbb{R}^n \setminus B(0,1)} = 0$. Consider the vector field $V : \mathbb{R}^n \to \mathbb{R}^n$, $V(x) = \varphi(x)c$. This is a smooth vector field with compact support. The flow generated by this vector field is complete, is the unique smooth map $F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\begin{aligned}F_0(x) &= x, \\ \frac{d}{dt}F_t(x) &= V(F_t(x)).\end{aligned}$$

We know that for each t the map $F_t : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism. Notice that $F_t(0) = ct$ is a solution for $-\frac{\delta}{\|c\|} < t < \frac{\delta}{\|c\|}$. That leads to $F_1(0) = c$. Therefore F_1 is a diffeomorphism of $\mathbb{R}^n$ bringing 0 to c. For each $x \in \mathbb{R}^n \setminus B(0,1)$, $F_t(x) = x$ is a solution, so F_t is the identity outside $B(0,1)$, in particular F_1 fixes the complement of $B(0,1)$. Note that F is a homotopy.

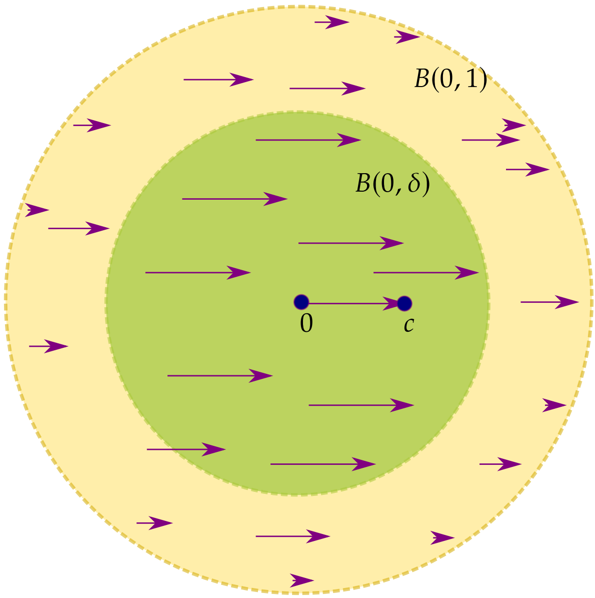


Figure 23.8: A flow pushing 0 to c, fixing the complement of the ball $B(0,1)$.

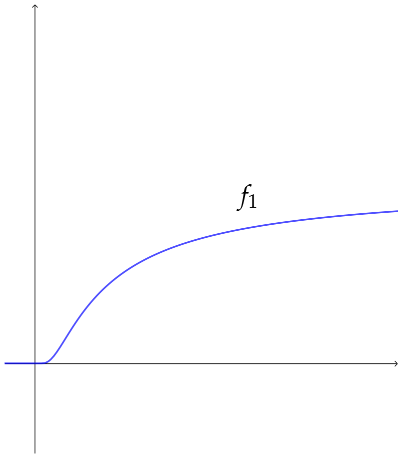
Now we prove the result globally using connectedness. Let M be a connected manifold. Let $x \in M$ and let S be the set of points of M such that there exists a diffeomorphism of M bringing x to y. The first part of this proof shows that S is an open set. If $z \in M$ is a limit point of S, take a sufficiently small neighborhood of z such that there is a diffeomorphism of M bringing any point in that neighborhood to z. That neighborhood contains a point y in S. There is diffeomorphism of M bringing x to y and a diffeomorphism bringing y to z, the composition brings x to z, thus $z \in S$. So S is both open and closed in M, implying $S = M$. Alternatively we can check that the relation that there is a diffeomorphism of M bringing x to y is an equivalence relation, and each equivalence class is open, so there is only one class. $\square$

Problems

23.9. ✓ Consider the following special function:

$$f_1(x) = \begin{cases} e^{-1/x}, & \text{if } x > 0 \\ 0, & \text{if } x \le 0. \end{cases}$$

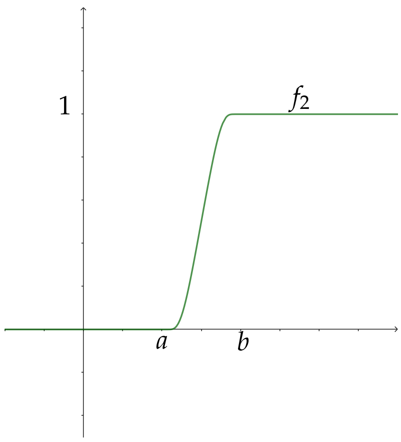
(a) Show that f_1 is smooth.



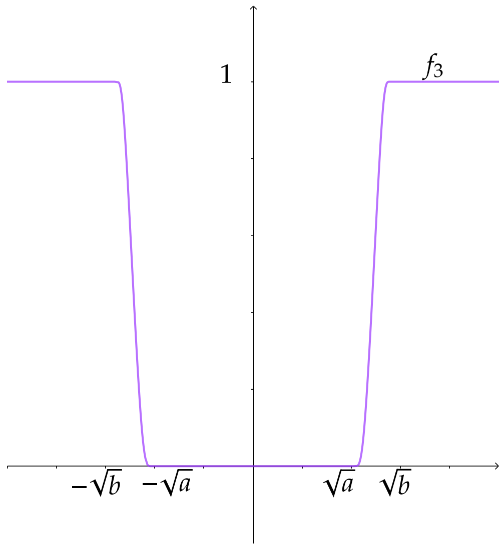
(b) With $0 < a < b$, the function

$$f_2(x) = \frac{f_1(x-a)}{f_1(x-a) + f_1(b-x)}$$

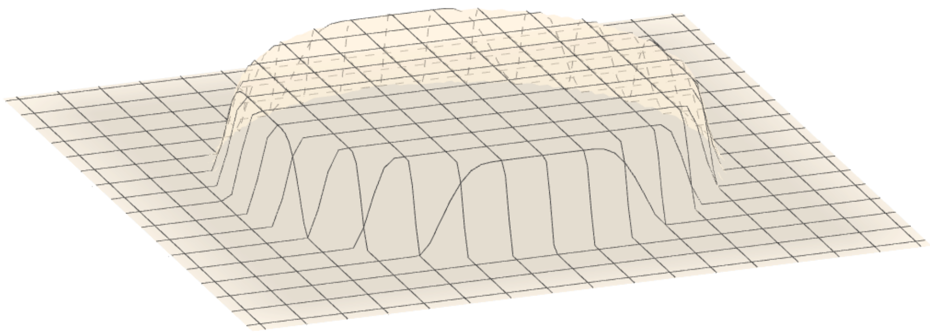
is smooth, $0 \leq f_2 \leq 1$, $f_2\big|_{(b,\infty)} \equiv 1$, $f_2\big|_{(-\infty,a)} \equiv 0$.



(c) With $0 < a < b$, the function $f_3(x) = f_2(x^2)$ is smooth, $0 \leq f_3 \leq 1$, $f_3|_{(-\sqrt{a},\sqrt{a})} = 0$, $f_3|_{\mathbb{R}\setminus(-\sqrt{b},\sqrt{b})} = 1$.



(d) In $\mathbb{R}^n$, with $0 < a < b$, let $f_4(x) = 1 - f_3(\|x\|^2)$. Show that f_4 is smooth, $0 \le f_4 \le 1$, $f_4\big|_{B(0,\sqrt[4]{a})} = 1$, $f_4\big|_{\mathbb{R}^n \setminus B(0,\sqrt[4]{b})} = 0$ This is a smooth function whose values are between 0 and 1, equal to 1 inside a smaller ball, equal to 0 outside of a bigger ball. This function can be used to connect different level sets smoothly.



This function is used often in Topology and Geometry where it can be called a "bump function", and in Analysis where it can be called a "test function" or a "mollifier".

23.10 (smooth Urysohn lemma). Let $A \subset U \subset \mathbb{R}^n$ where A is compact and U is open. We will show that there exists a smooth function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $0 \le \varphi \le 1$, $\varphi\,|_A = 1$, $\varphi\,\big|_{\mathbb{R}^n \setminus U} = 0$. Compare with the continuous Urysohn lemma in 9.7.

(a) Show that there exists a family of balls $(B(x_i, \epsilon_i))_{1 \le i \le k}$ such that $\bigcup_{i=1}^k B(x_i, \epsilon_i) \supset A$ and $\bigcup_{i=1}^k B(x_i, 2\epsilon_i) \subset U$.

(b) Let $\varphi_i : \mathbb{R}^n \to \mathbb{R}$ be smooth, $0 \le \varphi_i \le 1$, $\varphi_i\,\big|_{B(x_i,\epsilon_i)} \equiv 1$, $\varphi_i\,\big|_{\mathbb{R}^n \setminus B(x_i, 2\epsilon_i)} \equiv 0$. Let $\psi = \sum_{i=1}^k \varphi_i$. Show that $\psi\,|_A > 0$, $\psi\,\big|_{\mathbb{R}^n \setminus U} \equiv 0$.

(c) Let $c = \min_A \psi$. Let h be a smooth function such that $0 \le h \le 1$, $h\,\big|_{(-\infty,0]} \equiv 0$, $h\,\big|_{[c,\infty)} \equiv 1$. Let $\varphi = h \circ \psi$. Show that this is a function we are looking for.

23.11 (smooth Urysohn lemma for manifolds). Let M be a smooth m-dimensional manifold, $A \subset U \subset M$ where A is compact and U is open in M. We will show that there is a smooth function $\varphi : M \to \mathbb{R}$ such that $0 \le \varphi \le 1$, $\varphi\,|_A \equiv 1$, $\varphi\,\big|_{M \setminus U} \equiv 0$.

(a) Show that there exists a family of local parametrization $(\psi_i : \mathbb{R} \to M)_{1\le i\le k}$ and a family of balls $(B(x_i, \epsilon_i))_{1\le i\le k}$ in $\mathbb{R}^m$ such that $\bigcup_{i=1}^k \psi_i\,(B(x_i, \epsilon_i)) \supset A$ and $\psi_i\,(B(x_i, 2\epsilon_i)) \subset U$.

(b) Let $\varphi_i : \mathbb{R}^m \to \mathbb{R}$ be smooth, $0 \le \varphi_i \le 1$, $\varphi_i\big|_{B(x_i,\epsilon_i)} \equiv 1$, $\varphi_i\big|_{\mathbb{R}^m\setminus B(x_i,2\epsilon_i)} \equiv 0$. Let $\varphi_i' = \varphi_i \circ \psi_i^{-1}$ on $\psi_i(\mathbb{R}^m)$ and $\varphi_i'\big|_{M\setminus U} \equiv 0$. Let $\psi = \sum_{i=1}^k \varphi_i'$. Show that $\psi|_A > 0$, $\psi\big|_{M\setminus U} \equiv 0$. Then proceed as in 23.10.

23.12. Show that in 23.5 we can replace the assumption that the manifold M is compact by the assumption that the set $f^{-1}([a,b])$ is compact. The proof should be modified: we should use ψ which vanishes outside a compact neighborhood of $f^{-1}([a,b])$, that is, $\mathrm{supp}(\psi) = \mathrm{cl}(\{x \mid \psi(x) \neq 0\})$ is compact, while $\mathrm{supp}(\psi) \subset f^{-1}((h,k))$ (compare 6.33).

24 Boundary

The closed half-space $\mathbb{H}^m = \{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \mid x_m \geq 0\} \subset \mathbb{R}^m$, $m \geq 1$, whose topological boundary is $\partial\mathbb{H}^m = \{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \mid x_m = 0\}$ is our model for a manifold with boundary.

Definition. *A subspace M of $\mathbb{R}^k$ is called a* ***manifold with boundary***[1] *of dimension m if each point in M has a neighborhood diffeomorphic to either $\mathbb{R}^m$ or $\mathbb{H}^m$.*

The set of all points having neighborhoods diffeomorphic to $\mathbb{R}^m$ is called the ***interior*** *of M.*

If a point has a neighborhood diffeomorphic to $\mathbb{H}^m$ via a diffeomorphism that send the point to $\partial\mathbb{H}^m$ then it is called a boundary point of M. The set of all boundary points of M is called the ***boundary*** *of M, denoted by ∂M. Thus $x \in \partial M$ if and only if there is an open set U in M, $x \in U$, and there is $\varphi : U \to \mathbb{H}^m$ a diffeomorphism, such that $\varphi(x) \in \partial\mathbb{H}^m$.*

A point belongs to either the interior or the boundary, not both, since $\mathbb{H}^m$ is not diffeomorphic to $\mathbb{R}^m$, by smooth invariance of domain 19.3.

Example. The interval $[0, \infty)$ is a 1-dimensional manifold with boundary, the boundary is the 0-dimensional manifold $\{0\}$. The interval $[0, 1)$ is a 1-dimensional manifold with boundary, the boundary is the 0-dimensional manifold $\{0\}$. The interval $[0, 1]$ is a 1-dimensional manifold with boundary, the boundary is the 0-dimensional manifold $\{0, 1\}$.

24.1 Example. Let f be a smooth real function on $\mathbb{R}$. Consider the upper graph of f, that is, the region bounded below by the graph of f, $G = \{(x, y) \mid y \geq f(x), x \in \mathbb{R}\}$. We can derive a simple diffeomorphism from G to $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}$, namely $h(x, y) = (x, y - f(x))$. Thus G is a 2-dimensional manifold with boundary, and the boundary is exactly the graph of f.

[1] đa tạp có biên

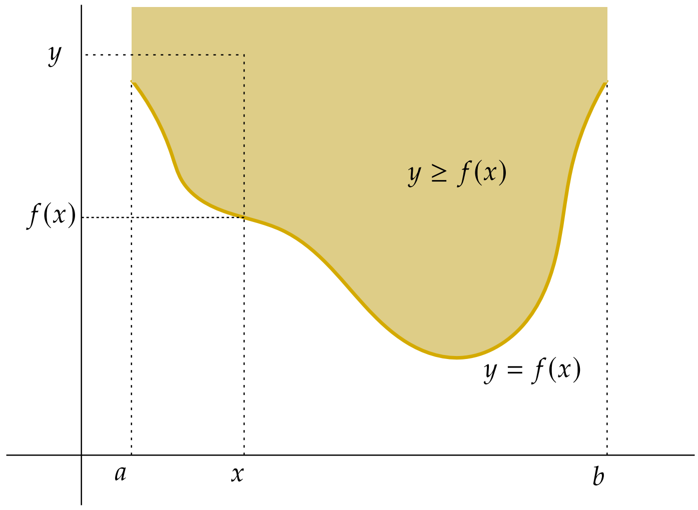


Figure 24.2: The uppergraph is a manifold with boundary.

Remark. We will follow a common convention that the term "manifold" is reserved for manifold without boundary. A manifold with boundary can have empty boundary, in which case it is a manifold. When there is a need to avoid confusion, we will say clearly as "let M be a manifold without boundary", or "let M be a manifold with or without boundary".

Proposition. *The interior of an m-manifold with boundary is an m-manifold without boundary. The boundary of an m-manifold with boundary is an $(m-1)$-manifold without boundary.*

Proof. The part about the interior is clear. Let us consider the part about the boundary.

Let M be an m-manifold and let $x \in \partial M$. Let φ be a diffeomorphism from a neighborhood U of x in M to $\mathbb{H}^m$. If $y \in U$ then $\varphi(y) \in \partial\mathbb{H}^m$ if and only if $y \in \partial M$. Thus the restriction $\varphi|_{U \cap \partial M}$ is a diffeomorphism from a neighborhood of x in ∂M to $\partial\mathbb{H}^m$, which is diffeomorphic to $\mathbb{R}^{m-1}$. □

The tangent space of a manifold with boundary M is defined as follows. It x is an interior point of M then TM_x is defined as before. If x is a boundary point then there is a parametrization $\varphi : \mathbb{H}^m \to M$, where $\varphi(0) = x$. Notice that by continuity φ has well-defined partial derivatives at 0 (see 19.14). This implies that the derivative $d\varphi_0 : \mathbb{R}^m \to \mathbb{R}^k$ is well-defined. Then TM_x is still defined as $d\varphi_0(\mathbb{R}^m)$. The chain rule for derivatives still holds.

Regular values for manifolds with boundaries

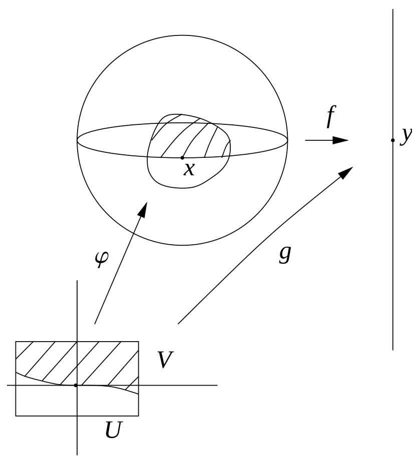
In Example 24.1, we see that the uppergraph of the smooth function f, the subset of the plane determined by the inequality $y \geq f(x)$, is a manifold with boundary. This is true more generally for essentially the same reason.

24.3 Theorem. *Let M be an m-dimensional manifold without boundary. Let $f : M \to \mathbb{R}$ be smooth and let y be a regular value of f. Then the* ***superlevel set*** *$\{x \in M \mid f(x) \geq y\} = f^{-1}([y, \infty))$ and the* ***sublevel set*** *$\{x \in M \mid f(x) \leq y\} = f^{-1}((-\infty, y])$ are m-dimensional manifolds with boundary the level set $f^{-1}(y)$.*

Proof. By noticing that $f(x) \geq y \iff -f(x) \leq -y$, we see that the case of superlevel set [1] implies the case of sublevel set [2]. We shall consider the case of superlevel set.

Let $N = f^{-1}([y, \infty))$. Since $f^{-1}((y, \infty))$ is an open subspace of M, it is an m-manifold without boundary.

Let $x \in f^{-1}(y)$. Let $\varphi : \mathbb{R}^m \to M$ be a parametrization of a neighborhood of x in M, with $\varphi(0) = x$. Let $g = f \circ \varphi$. As in the proof of 21.4, by the Implicit function theorem, there is an open ball U in $\mathbb{R}^{m-1}$ containing 0 and an open interval V in $\mathbb{R}$ containing 0 such that in $U \times V$ the set $g^{-1}(y)$ is a graph $\{(u, h(u)) \mid u \in U\}$ where h is smooth.



The set $(U \times V) \setminus g^{-1}(y)$ consists of two connected components, each component is mapped via g to either $(-\infty, y)$ or (y, ∞), but in fact exactly one of the two is mapped via g to (y, ∞), otherwise x will be a local extremum point of f, and so $df_x = 0$, violating the assumption. In order to be definitive, let us assume that $W = \{(u, v) \mid v \geq h(u)\}$ is mapped by g to $[y, \infty)$. Then $\varphi(W) = \varphi(U \times V) \cap f^{-1}([y, \infty))$ is a neighborhood of x in N parametrized by $\varphi|_W$. Notice that W is diffeomorphic to an open neighborhood of 0 in $\mathbb{H}^m$, as in 24.1, by considering the map $\psi(u, v) = (u, v - h(u))$ on $U \times V$. The map ψ is a smooth bijection on open subspaces of $\mathbb{R}^m$, its Jacobian is non-singular, hence it is a diffeomorphism. The restriction $\psi|_W$ is a diffeomorphism to $\psi(U \times V) \cap \mathbb{H}^m$. Thus x is a boundary point of N. □

Example. The closed disk D^n is an n-manifold with boundary S^{n-1}. Indeed, let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|^2$. Since 1 is a regular value of f, the sublevel set

[1] tập trên mức
[2] tập dưới mức

$D^n = f^{-1}((-\infty, 1])$ is an n-manifold with boundary $f^{-1}(1) = S^{n-1}$.

Example. Let f be the height function on S^2, $f(x, y, z) = z$. Since 0 is a regular value of f the set $f^{-1}([0, \infty))$, which is the upper hemisphere, is a 2-dimensional manifold with boundary $f^{-1}(0)$, which is the equator, diffeomorphic to S^1.

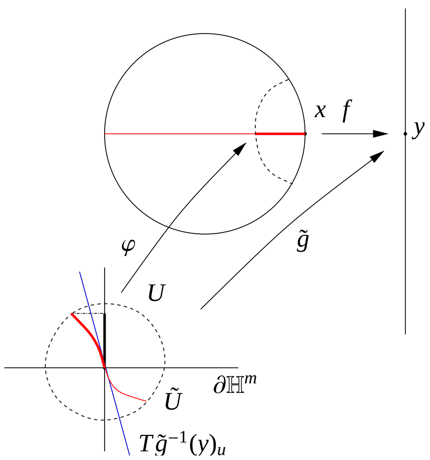
Notice in this example that the boundary of the upper hemisphere as a manifold, which is the equator, is different from the boundary of the upper hemisphere as a subset of the Euclidean space $\mathbb{R}^3$, which is the upper hemisphere itself. ***The boundary of a manifold is generally not the same as its topological boundary*** as a subset of $\mathbb{R}^k$.

The notion of critical point is defined for manifolds with boundaries exactly as for manifolds without boundaries.

24.4 Theorem. *Let M be an m-dimensional manifold with boundary, let N be an n-manifold with or without boundary, with $m > n$. Let $f : M \longrightarrow N$ be smooth. Suppose that $y \in N$ is a regular value of both f and $f|_{\partial M}$. Then $f^{-1}(y)$ is an $(m-n)$-manifold with boundary $\partial M \cap f^{-1}(y)$.*

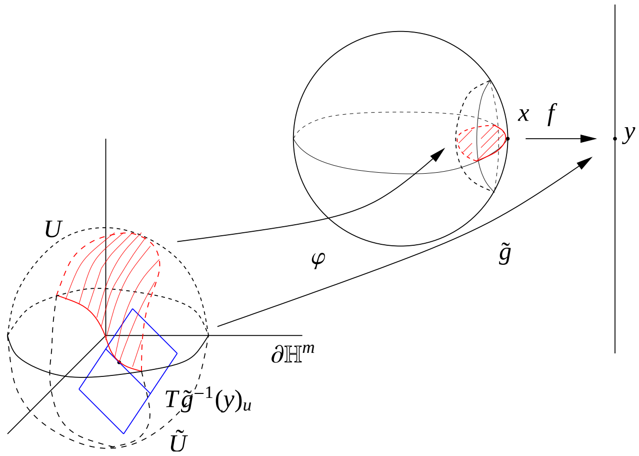
Proof. That $f^{-1}(y) \setminus \partial M$ is an $(m-n)$-manifold without boundary is already proved in 21.4.

We consider the crucial case of $x \in \partial M \cap f^{-1}(y)$.



Let $\varphi : \mathbb{H}^m \longrightarrow M$ be a parametrization of a neighborhood of x in M, with $\varphi(0) = x$. Let $g = f \circ \varphi$. The map g can be extended to $\tilde{g}$ defined on an open neighborhood $\tilde{U}$ of 0 in $\mathbb{R}^m$. Notice that $d\tilde{g}_0 = dg_0$, we have 0 is a regular point of $\tilde{g}$. By the Implicit function theorem, if we take a small enough neighborhood then $\tilde{g}^{-1}(y)$ is a graph of a function of $(m-n)$ variables so it is an $(m-n)$-manifold without boundary containing only regular points of $\tilde{g}$.

Let $p : \tilde{g}^{-1}(y) \longrightarrow \mathbb{R}$ be the projection to the last coordinate (the height function), namely $(x_1, x_2, \dots, x_m) \mapsto x_m$, then $g^{-1}(y) = p^{-1}([0, \infty))$. If we can show that 0 is a regular value of p then the desired result follows from 24.3



applied to $\tilde{g}^{-1}(y)$ and p. For each $u \in p^{-1}(0) \subset \partial\mathbb{H}^m$ the derivative dp_u is given by the vector e_m, so dp_u is onto if and only if the tangent space $T\tilde{g}^{-1}(y)_u$ is not contained in $\partial\mathbb{H}^m$.

Since $\tilde{g}$ is regular at u, the null space of $d\tilde{g}_u$ on $T\tilde{U}_u = \mathbb{R}^m$ is exactly $T\tilde{g}^{-1}(y)_u$, of dimension $m-n$. Since $\tilde{g}|_{\partial\mathbb{H}^m}$ is also regular at u, the null space of $d\tilde{g}_u$ restricted to $T(\partial\mathbb{H}^m)_u = \partial\mathbb{H}^m$ has dimension $(m-1)-n$. Thus $T\tilde{g}^{-1}(y)_u$ is not contained in $\partial\mathbb{H}^m$. □

Example. Let $f : D^2 \to \mathbb{R}$, $f(x,y) = y$. Then 0 is a regular value of f and $f^{-1}(0)$ is a 1-dimensional manifold with boundary on ∂D^2.

Sard theorem

We use the following result from Analysis:

Theorem (Sard theorem). *The set of critical values of a smooth map from $\mathbb{R}^m$ to $\mathbb{R}^n$ is of Lebesgue measure zero.*

For a proof see for instance [Mil97].

Sard theorem also holds for smooth functions from $\mathbb{H}^m$ to $\mathbb{R}^n$, see 24.14.

Since a set of Lebesgue measure zero must have empty interior, we have:

Corollary. *The set of regular values of a smooth map from $\mathbb{R}^m$ to $\mathbb{R}^n$ is dense in $\mathbb{R}^n$.*

Theorem. *If M and N are manifolds with boundaries and $f : M \to N$ is smooth then the set of all regular values of f is dense in N. Thus, **any smooth function has a regular value**.*

Proof. Consider any open subset V of N parametrized by $\psi : V' \to V$. Then $f^{-1}(V)$ is an open submanifold of M. We only need to prove that $f|_{f^{-1}(V)}$ has a regular value in V. Let C be the set of all critical points of $f|_{f^{-1}(V)}$.

We can cover $f^{-1}(V)$ (or any manifold) by a ***countable*** collection I of parametrized open neighborhoods. This is possible because a Euclidean space has a countable topological basis (see 2.24 and 2.20).

For each $U \in I$ we have a commutative diagram:

$$\begin{tikzcd} U \arrow[r, "f"] & V \\ U' \arrow[u, "\varphi_U"] \arrow[r, "g_U"] & V' \arrow[u, "\psi"'] \end{tikzcd}$$

where U' is an open subset of $\mathbb{H}^m$ and V' is an open subset of $\mathbb{H}^n$. From this diagram, x is a critical point of f in U if and only if $\varphi_U^{-1}(x)$ is a critical point of g_U. Thus the set of critical points of g_U is $\varphi_U^{-1}(C \cap U)$.

Now we write

$$f(C) = \bigcup_{U \in I} f(C \cap U) = \bigcup_{U \in I} \psi(g_U(\varphi^{-1}(C \cap U))) = \psi\left(\bigcup_{U \in I} g_U(\varphi^{-1}(C \cap U))\right).$$

By Sard Theorem the set $g_U(\varphi_U^{-1}(C \cap U))$ is of measure zero. This implies that the set $D = \bigcup_{U \in I} g_U(\varphi_U^{-1}(C \cap U))$ is of measure zero, since a countable union of sets of measure zero is a set of measure zero. As a consequence D must have empty topological interior.

Since ψ is a homeomorphism, $\psi(D) = f(C)$ must also have empty topological interior. Thus $f(C) \subsetneq V$, so there must be a regular value of f in V. □

Brouwer fixed point theorem

If $N \subset M$ and $f : M \to N$ such that $f|_N = \mathrm{id}_N$ then f is called a ***retraction*** from M to N and N is a ***retract*** of M.

24.5 Lemma. *Let M be a compact manifold with boundary. There is no smooth map $f : M \to \partial M$ such that $f|_{\partial M} = \mathrm{id}_{\partial M}$. In other words there is no smooth retraction from M to its boundary.*

Proof. Suppose that there is such a map f. Let y be a regular value of f. Since $f|_{\partial M}$ is the identity map, y is also a regular value of $f|_{\partial M}$. By Theorem 24.4 the inverse image $f^{-1}(y)$ is a 1-manifold with boundary $f^{-1}(y) \cap \partial M = \{y\}$. But a 1-manifold cannot have boundary consisting of exactly one point, see 24.6. □

24.6 Theorem (classification of compact one-dimensional manifolds). *A smooth compact connected one-dimensional manifold is diffeomorphic to either a circle, in which case it has no boundary, or a closed interval, in which case its boundary consists of two points.*

For a proof, see e.g. [Mil97] .

Repeating of the proof for the continuous Brouwer fixed point theorem in 17.7, we get:

24.7 Corollary (smooth Brouwer fixed point theorem). *A smooth map from the disk D^n to itself has a fixed point.*

Proof. Suppose that f does not have a fixed point, i.e. $f(x) \neq x$ for all $x \in D^n$. The straight line from $f(x)$ to x will intersect the boundary ∂D^n at a point $g(x)$. Then $g : D^n \to \partial D^n$ is a smooth function (see 24.17) which is the identity on ∂D^n. That is impossible, by 24.5. □

A proof for the continuous version of the theorem using the smooth version can be found in [Mil97].

Problems

24.8. Show that the inequality $x^6 + y^4 + z^2 \leq 1$ determines a 3-dimensional manifold with boundary in $\mathbb{R}^3$, and the boundary is given by $x^6 + y^4 + z^2 = 1$.

24.9. Show that the **solid torus** in $\mathbb{R}^3$, given by $\left(\sqrt{x^2+y^2}-b\right)^2 + z^2 \leq a^2$, is a 3-dimensional manifold with boundary, and the boundary is the torus surface, given by $\left(\sqrt{x^2+y^2}-b\right)^2 + z^2 = a^2$.

24.10. Show that the subspace $\mathbb{R}^m_+ = \{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \mid x_m > 0\}$ is diffeomorphic to $\mathbb{R}^m$.

24.11. Show that $U = \{(x, y) \mid x \in B(0, 1) \subset \mathbb{R}^{m-1}, y \geq 0\}$ is diffeomorphic to $\mathbb{H}^m$.

24.12. A simple regular path is a map $\gamma : [a, b] \to \mathbb{R}^m$ such that γ is injective, smooth, and $\gamma'(t) \neq 0$ for all $t \in [a, b]$. Show that the trace of a simple regular path is a smooth 1-dimensional manifold with boundary.

24.13. Suppose that M is a connected manifold without boundary. Show that $M \times [0, 1]$ is an $(n + 1)$-manifold with boundary. Show that the boundary of $M \times [0, 1]$ is $M \times \{0, 1\}$, consists of two connected components $M \times \{0\}$ and $M \times \{1\}$, each of which is diffeomorphic to M.

24.14. From the Sard theorem on $\mathbb{R}^m$, deduce that it also holds for smooth functions from $\mathbb{H}^m$ to $\mathbb{R}^n$.

24.15. Let $f : S^1 \to S^2$ be smooth.

(a) Find the regular points of f.

(b) Find the regular values of f.

(c) Show that f cannot be surjective.

(d) Deduce that that a smooth loop on the 2-dimensional sphere (i.e. a smooth map from S^1 to S^2) cannot cover the sphere.

(e) Generalize this result?

24.16. Show that there is no smooth surjective maps from $\mathbb{R}$ to $\mathbb{R}^n$ with $n > 1$. Deduce that there is no smooth space filling curve (compare the continuous case at 9.25).

24.17. Check that the function g in the proof of 24.7 is smooth.

24.18. Let M be a compact smooth manifold without boundary. Let $f : M \to \mathbb{R}$ be a smooth function. Assume that the interval $[a, b] \subset \mathbb{R}$ only contains regular values of f. Consider the flow ϕ in the proof of 23.5 on page 224.

(a) Show that $\forall x \in M, \forall t \in \mathbb{R}, 0 \leq \frac{d}{dt} f(\phi_t(x)) \leq 1$. Deduce that $\forall t \geq 0, f(x) \leq f(\phi_t(x)) \leq f(x) + t$.

(b) Show that if $x \in f^{-1}(a)$ then $\forall t \in [0, b-a], \frac{d}{dt} f(\phi_t(x)) = 1$. Deduce that $\forall t \in [0, b-a], f(\phi_t(x)) = a + t$.

(c) Check that the map

$$\begin{aligned} h = \phi\big|_{[0,b-a]\times f^{-1}(a)} : [0, b-a] \times f^{-1}(a) &\to f^{-1}([a, b]) \\ (t, x) &\mapsto \phi(t, x) = \phi_t(x) \end{aligned}$$

is well-defined.

(d) Check that the inverse map h^{-1} is given by $h^{-1}(y) = (f(y) - a, \phi_{a-f(y)}(y))$. Show that h is a diffeomorphism.

(e) Deduce that $f^{-1}([a, b])$ is a manifold with boundary diffeomorphic to $[0, b - a] \times f^{-1}(a)$.

(f) How is the compactness assumption used? Does the result remain correct without this assumption?

(g) How is the assumption that the interval $[a, b] \subset \mathbb{R}$ only contains regular values of f used? Does the result remain correct without this assumption?

24.19. Let M be a compact smooth manifold and $f : M \to \mathbb{R}$ be smooth. Show that if the interval $[a, b]$ only contains regular values of f then the sublevel sets $f^{-1}((-\infty, a])$ and $f^{-1}((-\infty, b])$ are diffeomorphic.

25 Orientation

Orientations on vector spaces

We have visualized orientations of $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$ as choices of one, two, or three vectors in certain orders. We also intuitively considered different choices as giving same orientations or opposite orientations (left – right, clockwise – counter-clockwise, left-hand – right-hand). Now we want to give precise meaning to this notion to be able to work in more complicated situations.

On a finite dimensional real vector space, an orientation is given by a vector basis. Two vector bases determine same orientations if the change of bases matrix has positive determinant. Being of the same orientation is an equivalence relation on the set of all vector bases. With this equivalence relation the set of all vector bases is divided into two equivalence classes. If we choose one of the two classes as the preferred one, then we say the vector space is ***oriented*** and the chosen equivalence class is called the ***orientation*** (or the positive orientation).

Thus any finite dimensional real vector space is ***orientable*** (i.e. can be oriented) with two possible orientations.

Example. The standard, canonical positive orientation of $\mathbb{R}^m$ is represented by the basis

$$(e_1 = (1,0,\dots,0), e_2 = (0,1,0,\dots,0), \dots, e_m = (0,\dots,0,1)).$$

Unless stated otherwise, $\mathbb{R}^m$ is oriented this way.

Let T be a linear isomorphism from an oriented finite dimensional real vector space V to an oriented finite dimensional real vector space W. Then T brings a basis of V to a basis of W. There are only two possibilities:

Lemma. *Either T brings any positive basis of V to a positive basis of W, or T brings any positive basis of V to a negative basis of W.*

Proof. Let $b = (b_1, \dots, b_n)$ and $b' = (b'_1, \dots, b'_n)$ be two bases of V of same orientations. That means the change of bases matrix $A = (a_{i,j})_{1\le i\le n, 1\le j\le n}$ has positive determinant. Since $T(b'_i) = T(\sum_{j=1}^n a_{j,i} b_j) = \sum_{j=1}^n a_{j,i} T(b_j)$, we see that the change of bases matrix from $(T(b_1), \dots, T(b_n))$ to $(T(b'_1), \dots, T(b'_n))$ is also A, so the two bases are of same orientations in W. $\square$

In the first case we say that T is ***orientation-preserving***, and in the second case we say that T is ***orientation-reversing***.

Orientations on manifolds

Roughly, a manifold is oriented if at each point an orientation for the tangent space is chosen and this orientation should be smoothly depended on the point.

Definition. *A smooth manifold M of dimension m, with or without boundary is said to be **oriented** if at each point x an orientation for the tangent space TM_x is chosen and at each point there exists a local parametrization φ such that the derivative $d\varphi_u : \mathbb{R}^m \to TM_{\varphi(u)}$ is orientation-preserving for all u.*

Thus the orientation of $TM_{\varphi(u)}$ is given by the basis $\left(\frac{\partial\varphi}{\partial u_1}(u), \frac{\partial\varphi}{\partial u_2}(u), \ldots, \frac{\partial\varphi}{\partial u_m}(u)\right)$ where $\frac{\partial\varphi}{\partial u_i}(u) = d\varphi_u(e_i)$. Roughly, local parametrizations bring the orientation of $\mathbb{R}^m$ to the manifold.

If a manifold is oriented then the set of orientations of its tangent spaces is called an ***orientation*** of the manifold and the the manifold is said to be ***orientable***.

Let $f : M \to N$ be a diffeomorphism between oriented manifolds. If df_x is orientation-preserving for all $x \in M$ then we say that f is an ***orientation-preserving diffeomorphism***. If df_x is orientation-reversing for all $x \in M$ then we say that f is an ***orientation-reversing diffeomorphism***.

Example. In $\mathbb{R}^2$ any translation is orientation-preserving, while any reflection over a coordinate axis orientation-reversing.

We can say that ***a manifold is oriented if each tangent space is oriented and each point has a neighborhood with an orientation-preserving parametrization***.

Another approach to orientation of manifold is to orient each parametrized neighborhood first then require that the orientations on overlapping neighborhoods agree. Concisely, suppose that $\varphi : \mathbb{R}^m \to M$ is a parametrization of a neighborhood of M and suppose that $\psi : \mathbb{R}^m \to M$ parametrizes an overlapping neighborhood. Since $d\psi_v = d(\psi \circ \varphi^{-1})_v \circ d\varphi_u$, the requirement for consistency of orientation is that the map $d(\psi \circ \varphi^{-1})_v$ must be orientation preserving on $\mathbb{R}^m$. Thus:

Proposition. *A manifold is orientable if and only if it has an atlas whose all change of coordinate functions are orientation-preserving.*

Example. If a manifold is parametrized by one parametrization, that is, it is covered by one local coordinate, then it is orientable, since we can take the unique parametrization to bring an orientation of $\mathbb{R}^m$ to the entire manifold. In particular, any open subset of $\mathbb{R}^k$ is an orientable manifold.

Example. The graph of a smooth function $f : D \to \mathbb{R}^l$, where $D \subset \mathbb{R}^k$ is an open set, is an orientable manifold, since this graph can be parametrized by a single parametrization, namely $x \mapsto (x, f(x))$.

Proposition. *A connected orientable manifold with or without boundary has exactly two orientations.*

Proof. Suppose the manifold M is orientable. There is an orientation o on M. Then $-o$ is a different orientation on M. Suppose that o_1 is an orientation on M, we show that o_1 is either o or $-o$.

If two orientations agrees at a point they must agree locally around that point. Indeed, from the definition there is a neighborhood V of x and a local coordinates $\varphi : V \to \mathbb{R}^m$ that brings the orientation o_1 to the standard orientation of $\mathbb{R}^m$, and a local coordinates $\psi : V \to \mathbb{R}^k$ that brings the orientation o to the standard orientation of $\mathbb{R}^m$. Assuming $\varphi(x) = \psi(x) = 0$, then $\det J(\psi^{-1} \circ \varphi)$ is smooth on $\mathbb{R}^m$ and is positive at 0, therefore it is always positive. That implies o_1 and o agree on V.

Let U be the set of all points x in M such that the orientation of TM_x with respect to o_1 is the same with the orientation of TM_x with respect to o. Then U is open in M. Similarly the complement $M \setminus U$ is also open. Since M is connected, either $U = M$ or $U = \emptyset$. ◻

25.1 Proposition. *If a $(k-1)$-dimensional manifold in $\mathbb{R}^k$ has a smooth normal unit vector field then it is orientable.*

For normal unit vector we are using the Euclidean inner product of $\mathbb{R}^k$, it means the vector is perpendicular to the tangent space. A submanifold of one dimension less is sometimes called a ***hypersurface*** (siêu phẳng). When a hypersurface has a smooth normal unit vector field it is sometimes called ***two-sided*** (hai phía).

Proof. Suppose M is a $(k-1)$-dimensional manifold in $\mathbb{R}^k$ having a smooth normal vector field $\boldsymbol{n}$. At each point $x \in M$ choose a linear basis $b(x) = (b_1(x), \dots, b_{k-1}(x))$ for T_xM such that $\tilde{b}(x) = (b_1(x), \dots, b_{k-1}(x), \boldsymbol{n}(x))$ is a linear basis in the positive orientation of $\mathbb{R}^k$, that is $\det \tilde{b}(x) > 0$. We check that this property depends only on the orientation class of $b(x)$. Suppose $c(x) = (c_1(x), \dots, c_{k-1}(x))$ is another basis of T_xM having the same orientation with $b(x)$, meaning the change of bases matrix $A(x)$ from $b(x)$ to $c(x)$ has $\det A(x) > 0$. Since $\boldsymbol{n}(x) \perp T_xM$ the vector $n(x)$ must be linearly independent from vectors in T_xM, therefore the change of bases matrix of $\mathbb{R}^k$ from $\tilde{b}(x) = (b_1(x), \dots, b_{k-1}(x), \boldsymbol{n}(x))$ to $\tilde{c}(x) = (c_1(x), \dots, c_{k-1}(x), \boldsymbol{n}(x))$ is

$$\begin{pmatrix} A(x) & 0 \\ 0 & 1 \end{pmatrix}$$

having determinant equal to $\det A(x) > 0$. Thus this rule using the unit normal determines an orientation for T_xM.

Let us check the smoothness of the orientation. Let $\varphi : \mathbb{R}^{k-1} \to M$ be a local parametrization of a neighborhood of $\varphi(u_0) = x_0$. At $x = \varphi(u)$ we have a basis $c(x) = \left(\frac{\partial\varphi}{\partial u_1}(u), \ldots, \frac{\partial\varphi}{\partial u_{k-1}}(u)\right)$ for T_xM. If $c(x_0)$ is not in the same orientation as $b(x_0)$ we can interchange two variables of φ to get a basis in the same orientation with $b(x_0)$, thus we can assume $c(x_0)$ has the same orientation with $b(x_0)$, that is $\det A(x_0) > 0$.

The function

$$u \mapsto \det\left(\tilde{c}(\varphi(u)\right) = \det\left(\frac{\partial\varphi}{\partial u_1}(u), \ldots, \frac{\partial\varphi}{\partial u_{k-1}}(u), n(\varphi(u))\right)$$

is smooth on $\mathbb{R}^{k-1}$. Since it is positive at u_0 it must always be positive.

From the above we get $\det\tilde{c}(x) = \det A(x)\det\tilde{b}(x)$. Since $\det\left(\tilde{b}((\varphi(u))\right)$ is always positive, we deduce that $\det A((\varphi(u))$ must always be positive. So $c((\varphi(u))$ is always in the same orientation with $b((\varphi(u))$. Thus φ is an orientation-preserving local parametrization, b gives an orientation for M, and M is orientable. □

Example. The sphere $S^n \subset \mathbb{R}^{n+1}$ has a unit normal vector field $F(x) = x$, therefore it is orientable.

Example. A two dimensional smooth manifold in $\mathbb{R}^3$ is called a (smooth) *surface*. By 25.1 above, any two-sided surface is orientable.

25.2 Proposition. *If $f : \mathbb{R}^k \to \mathbb{R}$ is smooth and a is a regular value of f then $f^{-1}(a)$ is an orientable manifold.*

Proof. Let $M = f^{-1}(a)$. If $x \in M$ then $\ker df_x = TM_x$, so the gradient vector $\nabla f(x)$ non-zero and is perpendicular to TM_x (the gradient vector is always perpendicular to the level set, see 22.1).

Thus ∇f is a nowhere zero smooth normal vector field and $\frac{\nabla f}{\|\nabla f\|}$ is a unit smooth normal vector field on$f^{-1}(a)$. We can now use 25.1. □

Example. The sphere and the torus are orientable, since they are level sets at regular values.

25.3 Proposition. *A surface is orientable if and only if it is two-sided.*

Proof. We check that if a surface S is orientable then it is two-sided. Choose an orientation for S. At each point $p \in S$ the tangent plane TS_p is oriented. Choose a positive vector basis (v_1, v_2) for TS_p. There is a unique vector $N_p \in \mathbb{R}^3$ such that $N_p \perp TS_p$, $\|N_p\| = 1$, and (v_1, v_2, N_p) is a positive vector basis of $\mathbb{R}^3$.

We check that N is smooth. There is a local parametrization $\varphi(u_1, u_2)$ such that the vector basis $(\varphi_{u_1}(u_1, u_2), \varphi_{u_2}(u_1, u_2))$ gives the positive orientation of $TS_{\varphi(u_1,u_2)}$. Then

$$N_{\varphi(u_1,u_2)} = \frac{\varphi_{u_1}(u_1, u_2) \times \varphi_{u_2}(u_1, u_2)}{||\varphi_{u_1}(u_1, u_2) \times \varphi_{u_2}(u_1, u_2)||}.$$

Thus N is smooth. □

It is possible to generalize the above result to hypersurfaces, see Problem 25.15.

Now we are able to prove a famous fact, that the Mobius surface (see 8.12) is not orientable.

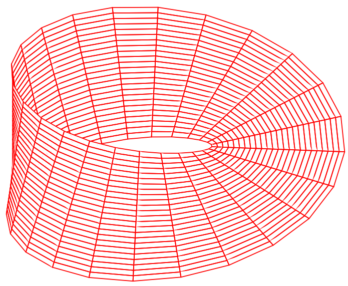


Figure 25.4: The Mobius surface is not orientable and is not two-sided.

Visually, if we pick a normal vector to the surface at a point in the center of the Mobius band, then move that normal vector smoothly along the center circle of the band. When we come back at the initial point after one loop, we realize that the normal vector is now in the opposite direction. That demonstrate that the Mobius surface is not two-sided. Similarly if we choose an orientation at a point then move that orientation continuously along the band then when we comeback the orientation has been switched.

Now we write this argument rigorously.

We define the Mobius surface (without boundary) as the image of the map (recall Fig. 8.15):

$$\begin{aligned}[0, 2\pi] \times (-1, 1) &\to \mathbb{R}^3 \\ (s, t) &\mapsto \left(\left(2 + t\cos\frac{s}{2}\right)\cos s, \left(2 + t\cos\frac{s}{2}\right)\sin s, t\sin\frac{s}{2}\right).\end{aligned}$$

25.5 Theorem. *The Mobius surface is not orientable.*

Proof. Let M be the Mobius surface and let

$$\begin{aligned}\varphi_1 : (0, 2\pi) \times (-1, 1) &\to M \\ (s, t) &\mapsto \left(\left(2 + t\cos\frac{s}{2}\right)\cos s, \left(2 + t\cos\frac{s}{2}\right)\sin s, t\sin\frac{s}{2}\right).\end{aligned}$$

We can check (though it is a bit tedious, as for polar coordinates) that this map is a local parametrization of M. This parametrization misses a subset of M, namely the interval $[1,3]$ on the x-axis. So we need one more parametrization to cover this part. We can take

$$\begin{aligned}\varphi_2 : (-\pi,\pi)\times(-1,1) &\to M\\ (s,t) &\mapsto \left(\left(2+t\cos\frac{s}{2}\right)\cos s, \left(2+t\cos\frac{s}{2}\right)\sin s, t\sin\frac{s}{2}\right).\end{aligned}$$

This φ_2 is given by the same formula as φ_1, but on a different domain. This parametrization misses the subset $\{-2\}\times\{0\}\times[-1,1]$ of M. Recall Fig. 8.15.

Suppose that M is orientable. Take an orientation for M. Then either φ_1 agrees with this orientation or disagrees with this orientation over the entire connected domain of φ_1. The same is true for φ_2. That implies that φ_1 and φ_2 either induce the same orientations over their entire domains, or they induces the opposite orientations over their domains.

Calculating directly, we get the normal vector given by φ_1 at the point $\varphi_1(s,0)$ on the center circle is:

$$(\varphi_1)_s \times (\varphi_1)_t(s,0) = \left(2\cos s\sin\frac{s}{2}, 2\sin s\sin\frac{s}{2}, -2\cos\frac{s}{2}\right).$$

The normal vector given by φ_2 at the point $\varphi_2(s,0)$ on the center circle is by the same formula:

$$(\varphi_2)_s \times (\varphi_2)_t(s,0) = \left(2\cos s\sin\frac{s}{2}, 2\sin s\sin\frac{s}{2}, -2\cos\frac{s}{2}\right).$$

At the point $(0,2,0) = \varphi_1(\frac{\pi}{2},0) = \varphi_2(\frac{\pi}{2},0)$ the two normal vectors agree, but at $(0,-2,0) = \varphi_1(\frac{3\pi}{2},0) = \varphi_2(-\frac{\pi}{2},0)$ they are opposite. Thus φ_1 and φ_2 do not give the same orientation, a contradiction. $\square$

Orientation on the boundary of an oriented manifold

The canonical orientation at every point of $\mathbb{H}^m$ is the canonical positive orientation of $\mathbb{R}^m$. The boundary $\partial\mathbb{H}^m$ is oriented as follows. At each point $x \in \partial\mathbb{H}^m$ the tangent space $T_x(\partial\mathbb{H}^m) = \mathbb{R}^{m-1}\times\{0\}$ is oriented by a basis $b = (b_1, b_2, \dots, b_{m-1})$ such that the ordered set

$$(-e_m, b_1, b_2, \dots, b_{m-1})$$

is a positive basis of $\mathbb{R}^m$. This is called the **outward normal first orientation of the boundary $\partial\mathbb{H}^m$**.

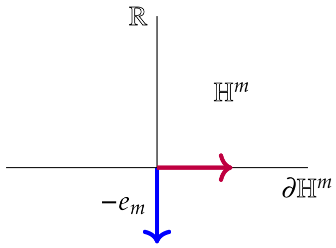


Figure 25.6: Outward normal first orientation of the boundary $\partial\mathbb{H}^m$.

Lemma. *This construction does not depend on the choice of representative basis b.*

Proof. Suppose $c = (c_1, \dots, c_{m-1})$ is another basis of $T_x(\partial\mathbb{H}^m)$ such that the change of bases matrix A from basis b to basis c has positive determinant. Then the change of bases matrix from basis $(-e_m, b_1, \dots, b_{m-1})$ to basis $(-e_m, c_1, \dots, c_{m-1})$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix},$$

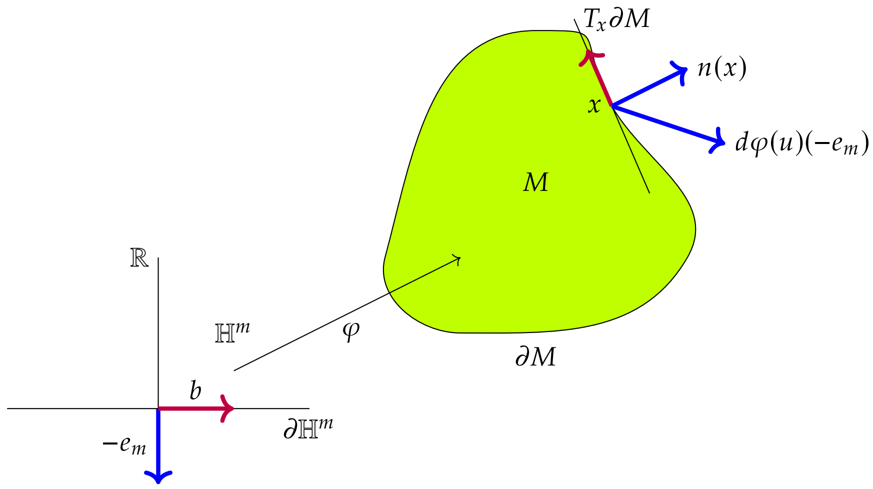
whose determinant is equal to $\det A > 0$. Thus c induced the same orientation on $T_x(\partial\mathbb{H}^m)$ as b. ☐

25.7 Example ($\partial\mathbb{H}^m = (-1)^m\mathbb{R}^{m-1}$). Take the canonical basis $(e_1, \dots, e_{m-1})$ of $\mathbb{R}^{m-1}$, and consider the basis $(i(e_1), \dots, i(e_{m-1}))$ of $\mathbb{R}^{m-1} \times \{0\}$ where i is the inclusion map $\mathbb{R}^{m-1} \hookrightarrow \mathbb{R}^{m-1} \times \{0\}$, $x \mapsto (x, 0)$. We have

$$\det(-e_m, i(e_1), \dots, i(e_{m-1})) = (-1)^m.$$

From this people often say that the boundary orientation of $\partial\mathbb{H}^m$ coincides with the canonical orientation of $\mathbb{R}^{m-1}$ if m is even and is opposite if m is odd, although this is an abuse of language.

Suppose that M is an oriented manifold with boundary. The boundary of M receives a special induced boundary orientation as follows. Suppose that $\varphi : \mathbb{H}^m \to M$ is an orientation-preserving local parametrization of M at the point $\varphi(u) = x \in \partial M$. Then $\varphi|_{\partial\mathbb{H}^m} : \partial\mathbb{H}^m \to \partial M$ is a local parametrization of ∂M at $\varphi(u)$. The derivative $d\varphi(u) : \mathbb{R}^m \to T_x M$ has a restriction $d\varphi(u) : T_x\partial\mathbb{H}^m \to T_x\partial M$. Let b be a basis of $T_x\partial\mathbb{H}^m$ giving the boundary orientation of $\partial\mathbb{H}^m$. The basis $d\varphi(u)(b)$ gives an orientation of $T_x\partial M$. This is called the **boundary orientation** of ∂M associated to the orientation of M.



Lemma. *This orientation does not depend on the choice of local parametrization.*

Proof. Suppose that $\psi : \mathbb{H}^m \longrightarrow M$ is another orientation-preserving local parametrization of M at the point $\psi(0) = x_0 \in \partial M$. Writing $\psi = \varphi \circ (\varphi^{-1} \circ \psi)$ we get $d\psi(v)(b) = d\varphi(u)(d(\varphi^{-1} \circ \psi)(v)(b))$. We need to check that the vector $d(\varphi^{-1} \circ \psi)(v)(b)$ is in the same orientation as b, i.e. $\varphi^{-1} \circ \psi$ is orientation-preserving on $\partial\mathbb{H}^m$. So this question is reduced to: Given an orientation-preserving $\varphi : \mathbb{H}^m \longrightarrow \mathbb{H}^m$, show that $\tilde{\varphi} = \varphi|_{\partial\mathbb{H}^m} : \partial\mathbb{H}^m \longrightarrow \partial\mathbb{H}^m$ is orientation-preserving. Let $u \in \partial\mathbb{H}^m$. The change of bases matrix from $(e_1, \ldots, e_m)$ to $(d\varphi(u)(e_1), \ldots, d\varphi(u)(e_m))$ is $J_\varphi(u)$, having positive determinant. Noting that $d\varphi(u)(e_i) = d\tilde{\varphi}(u)(e_i)$ for $1 \leq i \leq m-1$, we have

$$J_\varphi(u) = \begin{pmatrix} A & \\ 0 & \frac{\partial\varphi_m}{\partial u_m}(u) \end{pmatrix},$$

where A is the change of bases matrix from $(e_1, \ldots, e_{m-1})$ to $(d\tilde{\varphi}(u)(e_1), \ldots, d\tilde{\varphi}(u)(e_{m-1}))$. Now we have $\det J_\varphi(u) = \frac{\partial\varphi_m}{\partial u_m}(u) \det A > 0$. We write

$$\frac{\partial\varphi_m}{\partial u_m}(u) = \lim_{t\to 0^+} \frac{\varphi_m(u + te_m) - \varphi_m(u)}{t} = \lim_{t\to 0^+} \frac{\varphi_m(u + te_m)}{t},$$

noticing that since φ brings u to a point in $\partial\mathbb{H}^m$ we must have $\varphi_m(u) = 0$. Since $u + te_m \in \operatorname{int}(\mathbb{H}^m)$ when $t > 0$ we have $\varphi(u + te_m) \in \operatorname{int}(\mathbb{H}^m)$, so $\varphi_m(u + te_m) > 0$. Thus $\frac{\partial\varphi_m}{\partial u_m}(u) \geq 0$. This implies actually that $\frac{\partial\varphi_m}{\partial u_m}(u) > 0$, and that $\det A > 0$. Thus $d\tilde{\varphi}(u)$ is orientation-preserving.

Notice that, since $\frac{\partial\varphi_m}{\partial u_m}(u) = d\varphi(u)(e_m) \cdot e_m$, the problem has been essentially reduced to checking that the derivative $d\varphi(u)$ brings e_m to an inward pointing vector, i.e. $d\varphi(u)(e_m) \cdot e_m > 0$. □

We say that a vector $v \in \mathbb{R}^m$ is called **inward pointing** relative to $\mathbb{H}^m$ if $v \cdot e_m > 0$, and is **outward pointing** relative to $\mathbb{H}^m$ if $v \cdot e_m < 0$. At $x \in \partial M$, a vector $w \in T_xM$ is called **outward pointing** if there is an orientation-preserving local parametrization φ such that $x = \varphi(u)$ and $w = d\varphi(u)(v)$ where $v \in \mathbb{R}^m$ is outward pointing relative to $\mathbb{H}^m$. As observed in the last paragraph of the above proof, this notion does not depend on the choice of parametrization. Thus the orientation of the boundary can be called the **outward pointing first orientation of the boundary**.

In the Euclidean inner product of $\mathbb{R}^k$, there is a unique unit vector $n(x) \in T_xM$ such that $n(x) \perp T_x\partial M$ and $n(x)$ is outward pointing. This vector $n(x)$ is called the **outward unit normal vector of the boundary**.

With this special outward pointing vector, we can describe the boundary orientation of ∂M as the one represented by a basis $c(x)$ of $T_x\partial M$ such that $(n(x), c(x))$ represents the orientation of T_xM, and so the orientation is also called the **outward normal first orientation of the boundary**.

Example. The circle S^1 receives the induced outward normal first orientation as the boundary of the closed disk D^2. As such the circle is oriented counter-clockwise.

Example. The sphere S^2 receives the induced outward normal first orientation as the boundary of the closed ball D^3. As such the sphere is oriented counter-clockwise viewed from outside the ball.

Problems

25.8. Show that two diffeomorphic manifolds are are either both orientable or both un-orientable. For example, this implies that the Mobius band and the cylinder are not diffeomorphic.

25.9. Suppose that $f : M \to N$ is a diffeomorphism of connected oriented manifolds with boundaries. Show that if there is an x such that $df_x : TM_x \to TN_{f(x)}$ is orientation-preserving then f is orientation-preserving.

25.10. Let $f : \mathbb{R}^k \to \mathbb{R}^l$ be smooth and let a be a regular value of f. Show that $f^{-1}(a)$ is an orientable manifold.

25.11. Consider the map $-\mathrm{id} : S^n \to S^n$ with $x \mapsto -x$. Show that $-\mathrm{id}$ is orientation-preserving if and only if n is odd.

25.12. Discuss the boundary orientation for a 1-dimensional manifold.

25.13. Discuss the boundary orientation for the cylinder $S^1 \times [0, 1]$.

25.14. ✓ Let M be an orientable manifold without boundary. From 24.13, $M \times [0, 1]$ is a manifold with boundary $(M \times \{0\}) \cup (M \times \{1\})$.

(a) Show that $M \times [0, 1]$ is orientable.

(b) Show that with boundary orientation of $M \times [0, 1]$, exactly one of the two boundary components $M \times \{0\}$ and $M \times \{1\}$ has same orientation as that of M (meaning for example the map $(x, 0) \mapsto x$ is orientation-preserving) while the other has opposite orientation to that of M.

25.15. Generalize Proposition 25.3 to hypersurfaces.

25.16. * A **smooth isotopy** is an isotopy by diffeomorphisms which is smooth (recall the notion of isotopy in 12.12). Namely two diffeomorphisms f and g from smooth manifold M to smooth manifold N are **smoothly isotopic** if there is a smooth map, called a smooth isotopy

$$\begin{aligned} F : M \times [0,1] &\to N \\ (x,t) &\mapsto F(x,t) \end{aligned}$$

such that $F(x,0) = f(x)$ and $F(x,1) = g(x)$ for all $x \in M$, and for each $t \in [0,1]$ the map $F(\cdot,t)$ is a diffeomorphism.

Show that at each point two isotopic diffeomorphisms must be both orientation-preserving or both orientation-reversing.

26 Topological degrees of maps

Let $f : M \to N$ and let y be a regular value of f. If $\dim M = \dim N$ then at any $x \in f^{-1}(y)$ the linear map df_x is an isomorphism. If M and N are oriented then df_x is either orientation-preserving or orientation-reversing. We can assign the number 1 to the first case and the number -1 to the second case, and if the set $f^{-1}(y)$ is finite then we count the sum. Under some conditions this sum turns out to be independent of the choice of y, and is called the degree of f.

Let us now go into more details. Let M and N be boundaryless, oriented manifolds of the same dimensions m. Let $f : M \to N$ be smooth. If x is a regular point of f then df_x is an isomorphism from TM_x to $TN_{f(x)}$, let $\operatorname{sign}(df_x) = 1$ if df_x preserves orientations and $\operatorname{sign}(df_x) = -1$ otherwise. For any regular value y of f, let

$$\deg(f, y) = \sum_{x \in f^{-1}(y)} \operatorname{sign}(df_x).$$

This sum is defined when the set $f^{-1}(y)$ is finite, such as when M is compact (see 21.3), if $f^{-1}(y)$ is empty the sum is assigned to be 0. This number $\deg(f, y)$ is called the **Brouwer degree** [1] or **topological degree** of the map f with respect to the regular value y.

From the Inverse function theorem 21.23, each regular value y has a neighborhood V and each preimage x of y has a neighborhood U_x on which f is a diffeomorphism onto V, either preserving or reversing orientation. Therefore we can interpret that $\deg(f, y)$ ***counts the algebraic number of times the function f covers the value y***.

Example. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Then $\deg(f, 1) = 0$. This could be explained geometrically from the graph of f, as f covers the value 1 twice in opposite directions at $x = -1$ and $x = 1$. Indeed, df_x is orientation-preserving if and only if $f'(x) = 2x > 0$. Thus $\operatorname{sign} df_{-1} = -1$ while $\operatorname{sign} df_1 = 1$.

Example. Consider $f(x) = x^3 - x$ with the regular value 0. From the graph of f we see that f covers the value 0 three times in positive direction at $x = -1$ and $x = 1$ and negative direction at $x = 0$, therefore we see right away that $\deg(f, 0) = 1$. Indeed, df_x is orientation-preserving if and only if $f'(x) = 3x^2 - 1 > 0$. Thus $\operatorname{sign} df_{-1} = \operatorname{sign} df_1 = 1$ while $\operatorname{sign} df_0 = -1$.

If we consider the regular value 6 then visually f covers this value only once in positive direction at $x = 2$, thus $\deg(f, 1) = 1$. Indeed $\operatorname{sign} df_2 = 1$.

Two smooth maps are said to be **smoothly homotopic** if they are homotopic via a smooth homotopy, compare with continuous homotopy in Section 12,

[1] L. E. J. Brouwer (1881–1966) made important contributions in the early development of topology.

and see 26.25.

The main result of this section is that under some conditions ***the topological degree does not depend on the choice of the regular value and is invariant under homotopy of the map***.

26.1 Theorem (degree is a homotopy invariant). *Let M and N be boundaryless, oriented manifolds of .same dimensions, let M be compact and N be connected. Then the degree of a map from M to N does not depend on the choice of regular values and is invariant under smooth homotopy.*

Therefore under those assumptions we can write $\deg(f)$ instead of $\deg(f, y)$.

Example. On a compact, connected, oriented, boundaryless manifold:

- The degree of an identity map is 1.
- The degree of a constant map is 0.
- The degree of a diffeomorphism is ± 1.

Example. On a compact, connected, oriented, boundaryless manifold, the degree of the identity map is 1 while the degree of a constant map is 0, therefore the identity map is not homotopic to a constant map.

Example. A rotation around the origin of $\mathbb{R}^2$ restricted to the circle S^1 is a diffeomorphism of S^1. It is apparent that it is orientation-preserving on S^1. Thus its degree is 1.

The reflection about the x-axis of $\mathbb{R}^2$ restricted to the circle S^1 is a diffeomorphism of S^1. It is apparent that it is orientation-reversing on S^1. Thus its degree is -1.

Example. Consider the reflection about the xy-plane in $\mathbb{R}^3$, restricted to the sphere S^2:

$$\begin{aligned} f : S^2 &\to S^2 \\ (x, y, z) &\mapsto (x, y, -z). \end{aligned}$$

It is probably no longer apparent whether this map is orientation-preserving or orientation-reversing. Let us compute.

Consider the value $(1, 0, 0) \in S^2$, the only preimage is the point $(1, 0, 0)$ itself. The tangent space $TS^2_{(1,0,0)}$ is perpendicular to the point $(1, 0, 0)$, thus $TS^2_{(1,0,0)} = \{0\} \times \mathbb{R}^2$.

Since f is the restriction of the linear map on $\mathbb{R}^3$ given by the same formula, the derivative $df_{(x,y,z)}$ is also given by the same formula, that is, $df_{(x,y,z)}(u, v, w) = (u, v, -w)$.

We have $df_{(1,0,0)}(0, 1, 0) = (0, 1, 0)$ and $df_{(1,0,0)}(0, 0, 1) = (0, 0, -1)$. Clearly the basis $((0, 1, 0), (0, 0, 1))$ and the basis $((0, 1, 0), (0, 0, -1))$ represent opposite orientations of $TS^2_{(1,0,0)}$. Thus $df_{(1,0,0)}$ is orientation-reversing. We conclude that $\deg f = -1$.

26.2 Example. Let $f : S^1 \to S^1$, $f(z) = z^2$, using complex number notation as in 21.17.

In the following we use both notations. Consider a value $1 \in \mathbb{C}$,

$$f^{-1}(1) = \{z \in \mathbb{C} \mid z^2 = 1\} = \{\pm 1\} \subset \mathbb{C},$$

in other words for value $(1,0) \in S^1$

$$f^{-1}((1,0)) = \{(1,0), (-1,0)\} \subset S^1.$$

Take the counter-clockwise orientation of S^1. To determine the derivative of f, it is convenient to take a path $\gamma(t) = (\cos t, \sin t)$ on S^1, for which $\gamma'(t) = (-\sin t, \cos t)$, $\gamma(0) = (1,0)$, $\gamma'(0) = (0,1)$. Then

$$\begin{aligned} df_{(1,0)}((0,1)) &= df_{\gamma(0)}(\gamma'(0)) = (f \circ \gamma)'(0) \\ &= \frac{d}{dt}(\gamma(t))^2\bigg|_{t=0} = \frac{d}{dt}(e^{it})^2\bigg|_{t=0} \\ &= 2ie^{i2t}\big|_{t=0} = 2i = (0,2). \end{aligned}$$

Thus f is orientation-preserving at $(1,0)$.

Similarly, at $(-1,0) = \gamma(\pi)$, we have $\gamma'(\pi) = (0,-1)$, and

$$\begin{aligned} df_{(-1,0)}((0,-1)) &= df_{\gamma(0)}(\gamma'(\pi)) = (f \circ \gamma)'(\pi) \\ &= \frac{d}{dt}(\gamma(t))^2\bigg|_{t=\pi} = \frac{d}{dt}(e^{it})^2\bigg|_{t=\pi} \\ &= 2ie^{i2t}\big|_{t=\pi} = 2i = (0,2). \end{aligned}$$

Thus f is also orientation-preserving at $(-1,0)$. Therefore $\deg f = 1 + 1 = 2$.

26.3 Proposition. *Let M, N, P be compact, connected, oriented, boundaryless manifolds of same dimensions. Let $M \xrightarrow{f} N \xrightarrow{g} P$. Then* $\deg(g \circ f) = \deg(f)\deg(g)$.

Proof. This is a simple calculation using the chain rule for derivative:

$$\begin{aligned}
\deg(g\circ f, z) &= \sum_{x\in(g\circ f)^{-1}(z)} \operatorname{sign} d(g\circ f)_x \\
&= \sum_{x\in(g\circ f)^{-1}(z)} \operatorname{sign}\left(dg_{f(x)}\circ df_x\right) \\
&= \sum_{x\in(g\circ f)^{-1}(z)} \left(\operatorname{sign} dg_{f(x)}\right)\left(\operatorname{sign} df_x\right) \\
&= \sum_{y\in g^{-1}(z), x\in f^{-1}(y)} \left(\operatorname{sign} dg_y\right)\left(\operatorname{sign} df_x\right) \\
&= \sum_{y\in g^{-1}(z)} \operatorname{sign} dg_y \sum_{x\in f^{-1}(y)} \operatorname{sign} df_x \\
&= \sum_{y\in g^{-1}(z)} \left(\operatorname{sign} dg_y\right)\deg(f, y) \\
&= \deg(f)\sum_{y\in g^{-1}(z)} \operatorname{sign} dg_y = \deg(f)\deg(g, z) = \deg(f)\deg(g).
\end{aligned}$$

In the later part we use the independence of degree from the choice of regular value, Theorem 26.1. □

Example. Consider the antipodal map

$$\begin{aligned}
f : S^2 &\to S^2 \\
(x, y, z) &\mapsto (-x, -y, -z).
\end{aligned}$$

Since this map is a composition of three reflections about coordinate planes, each of which is of degree -1, by 26.3, $\deg f = (-1)(-1)(-1) = -1$.

Proof of homotopy invariance of degree

Here is a sketch of ideas for a proof of the main result, Theorem 26.1. See Fig. 26.4. Suppose $f, g : M \to N$ are homotopic via a smooth map $H : M\times[0,1] \to N$. Let $\partial H = H|_{\partial(M\times[0,1])}$. Consider the case y is a common regular value of f, g, H. Then $H^{-1}(y)$ is a 1-dimensional compact manifold whose boundary is $H^{-1}(y)\cap\partial(M\times[0,1]) = (\partial H)^{-1}(y) = f^{-1}(y)\cup g^{-1}(y)$. Consideration of orientations leads to the observation that ∂H at the two endpoints of an arc component of $H^{-1}(y)$ have opposite orientation preserving properties, thus $\deg\partial H = 0$. This implies $\deg f - \deg g = 0$, since $M\times\{0\}$ has opposite orientation to $M\times\{1\}$.

Now we work on details.

Let $X = M\times[0,1]$, a smooth manifold of 1 dimension higher than M, whose boundary is $\partial X = (M\times\{0\})\cup(M\times\{1\})$. From 25.14, with boundary orientation of ∂X, exactly one of the two boundary components $M\times\{0\}$ and

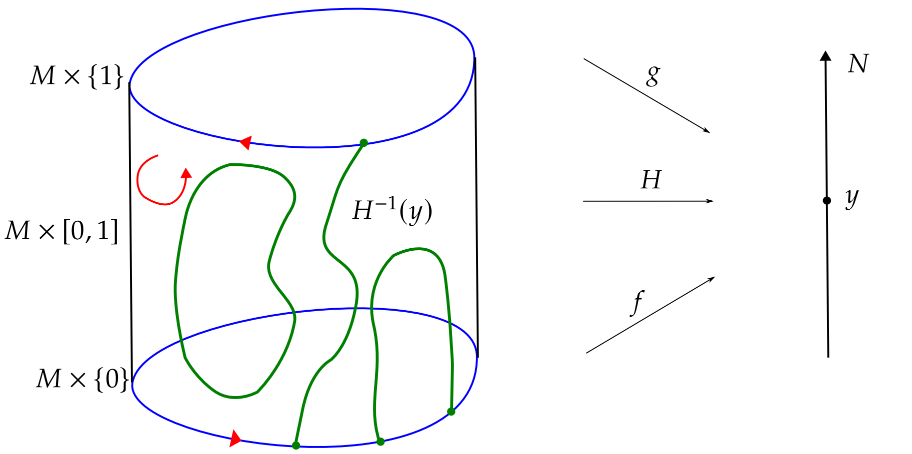


Figure 26.4: Ideas for a proof of homotopy invariance of degree.

$M \times \{1\}$ has the same orientation as the orientation of M while the other has opposite orientation to the orientation of M.

Let y be a common regular value of f and g.

Note that the map ∂H consists of f and g, precisely, for $x \in M$, we have $\partial H(x,0) = f(x)$ and $\partial H(x,1) = g(x)$.

(a) Assume that y is also a regular value of H. Then $H^{-1}(y)$ is a manifold of dimension 1 whose boundary is $H^{-1}(y) \cap (\partial X) = (\partial H)^{-1}(y)$, by Theorem 24.3. By the classification of one-dimensional manifolds 24.6, $H^{-1}(y)$ is the disjoint union of arcs and circles. Let A be a component that intersects ∂X. Then A is an arc with boundary $\{a,b\} \subset \partial X$.

Lemma. $\operatorname{sign} d(\partial H)_a = -\operatorname{sign} d(\partial H)_b$.

Proof. Choose an orientation for X. Choose an orientation for A as follows. For $x \in A$, let $(v_1, v_2, v_3, \ldots, v_{n+1})(x)$ be a positive oriented basis for TX_x such that $v_1(x) \in TA_x$ and $(dH_x(v_2(x)), \ldots, dH_x(v_{n+1}(x)))$ gives the positive orientation of TN_y. Since TA_x is the kernel of $dH_x : TX_x \to TN_y$ (see 21.22), the map dH_x is bijective on the vector subspace generated by $v_2(x), v_3(x), \ldots, v_{n+1}(x)$. Let the positive direction of TA_x be given by $v_1(x)$. See Fig. 26.5.

The idea is that since v_1 gives a smooth orientation of A, precisely one of the two vectors $v_1(a)$ and $v_1(b)$ is inward pointing while the other is outward pointing with respect to X, contradicting that $(v_1, v_2, v_3, \ldots, v_{n+1})$ belongs to a smooth orientation for X, which demands that ∂X has the outward pointing first orientation.

First we check that v_1 gives a smooth orientation for A. Let φ be a local parametrization of a neighborhood of $x = \varphi(u) \in A$ in X. Since y is a regular value of $H \circ \varphi$, the preimage $(H \circ \varphi)^{-1}(y) = \varphi^{-1}(A)$ is a 1-dimensional manifold. We can assume that in a sufficiently small neighborhood $\varphi^{-1}(A)$ is the graph of a function of the variable u_1. Indeed, since the Jacobian matrix $J(H \circ \varphi)_u$ has rank n, we can arrange the order of the variables of φ so that the last n columns

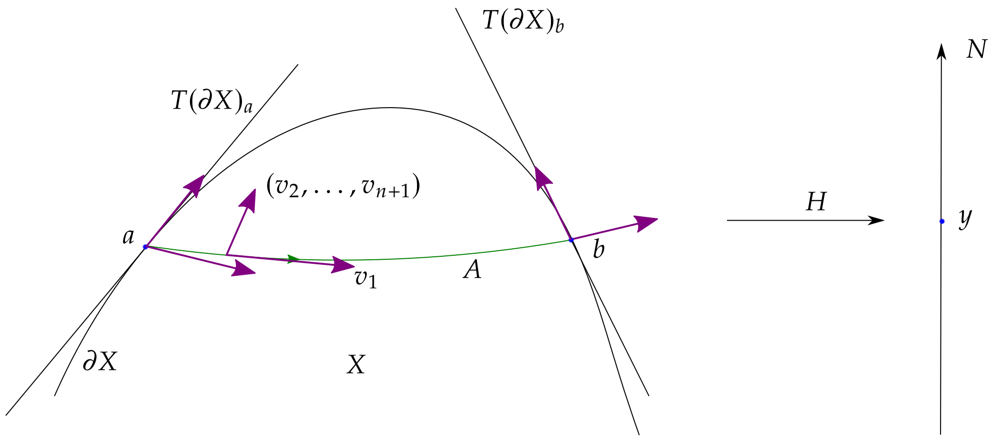


Figure 26.5: If the orientation of A is inward pointing with respect to X at a then it is outward pointing at b.

of this matrix are independent, then apply the Implicit function theorem to obtain a neighborhood of u in $\mathbb{R}^{n+1}$ in which the equation $(H \circ \varphi)(u) = y$ determines the graph $\{(u_1, g(u_1) \mid u_1 \in (c, d) \subset \mathbb{R}\}$ where $g : (c, d) \to \mathbb{R}^n$ is smooth (compare the proofs of 21.1 and 24.3). Bring the tangent vector $(1, g'(u_1))$ of this graph to the tangent vector $v_1(\varphi(u)) = d\varphi_u(1, g'(u_1))$ of A, and take $v_i(\varphi(u)) = d\varphi_u(e_i) = \frac{\partial \varphi}{\partial u_i}(u)$ for $2 \leq i \leq n+1$. These $(n+1)$ vectors are linearly independent in $TX_{\varphi(u)}$ since the vectors $(1, g'(u_1))$ and e_i for $i \geq 2$ are linearly independent in $\mathbb{R}^{n+1}$. If $(dH_x(v_2(x)), \ldots, dH_x(v_{n+1}(x)))$ does not give the positive orientation of TN_y we replace v_2 by $-v_2$. If $(v_1, v_2, v_3, \ldots, v_{n+1})$ does not give the positive orientation of X we replace v_1 by $-v_1$. Then we obtain $(v_1, v_2, v_3, \ldots, v_{n+1})$ as the desired basis where v_1 is smooth on A.

Now we check that one of the two vectors $v_1(a)$ and $v_1(b)$ is inward pointing and the other is outward pointing with respect to X.

First we claim that $v_1(a) \notin T(\partial X)_a$. Indeed, since a is a regular point of ∂H, the map $d(\partial H)_a : T(\partial X)_a \to TN_y$ is bijective, thus if $v_1(a) \in T(\partial X)_a$ then $d(\partial H)_a(v_1(a)) \neq 0$. On the other hand $v_1(a) \in TA_a = \ker(dH_a)$, so $dH_a(v_1(a)) = 0$, hence $d(\partial H)_a(v_1(a)) = 0$, a contradiction.

Figure 26.6: A path on $\mathbb{H}^{n+1}$ whose endpoints are on $\partial\mathbb{H}^{n+1}$ must be inward pointing at the start and outward pointing at the end.

Let A be oriented as an arc going from a to b, parametrized by a smooth regular path γ with $\gamma(0) = a$, $\gamma(1) = b$, and $\forall t \in [0, 1]$, $\gamma'(t) \neq 0$, see Fig 26.6. In a local coordinate $\varphi : \mathbb{H}^{n+1} \to X$ at $a = \varphi(0)$, consider the corresponding path $\alpha = \varphi^{-1} \circ \gamma$ in $\mathbb{H}^{n+1}$. Since $\alpha_{n+1}(0) = 0$ and $\alpha_{n+1}(t) \geq 0$ for $t > 0$, we have $\alpha'_{n+1}(0) \geq 0$. Further $\alpha'_{n+1}(0) \neq 0$, otherwise $\alpha'(0) \in \partial\mathbb{H}^{n+1}$ implying $v_1(a) = \gamma'(0) = d\varphi_0(\alpha'(0)) \in T(\partial X)_a$, a contradiction. Thus $\alpha'_{n+1}(0) > 0$, hence $\alpha'(0)$ points inward with respect to $\mathbb{H}^{n+1}$, and so $\gamma'(0) = v_1(a)$ points inward with respect to X.

Argue similarly we find that $\gamma'(1) = v_1(b)$ points outward. □

With the preceding lemma, taking sum over all arc components of $H^{-1}(y)$ gives us $\deg(\partial H, y) = 0$. Since $\deg(\partial H, y) = \pm(\deg(f, y) - \deg(g, y))$, we have $\deg(f, y) = \deg(g, y)$.

(b) Now suppose that y is not a regular value of H. We show that there is a neighborhood U of y such that for any $z \in U$ we have $\deg(\partial H, z) = \deg(\partial H, y)$.

Since y is a regular value of ∂H, for each $x \in (\partial H)^{-1}(y)$ there is an open connected neighborhood U_x of x on which ∂H is a diffeomorphism. By compactness of M the discrete set $(\partial H)^{-1}(y)$ is finite (compare 21.18). Let

$$V = \left[\bigcap_{x \in (\partial H)^{-1}(y)} f(U_x) \right] \setminus \partial H \left(M \setminus \bigcup_{x \in (\partial H)^{-1}(y)} U_x \right)$$

(compare 21.19). The set V is open, containing y. For any $z \in V$, we have $z \notin \partial H \left(M \setminus \bigcup_{x \in (\partial H)^{-1}(y)} U_x \right)$, hence $(\partial H)^{-1}(z) \subset \bigcup_{x \in (\partial H)^{-1}(y)} U_x$. For each $x \in (\partial H)^{-1}(y)$, since $V \subset \partial H(U_x)$ there is a unique $u \in (\partial H)^{-1}(z) \cap U_x$. For $u \in U_x$, since U_x is connected, the maps df_u and df_x are either both orientation-preserving or both orientation-reversing. It follows that $\deg(\partial H, z) = \deg(\partial H, y)$.

By Sard theorem, ∂H must have a regular value z in U. Then $\deg(\partial H, z) = 0$ by (a), thus $\deg(\partial H, y) = 0$. This implies $\deg(f, y) = \deg(g, y)$, as in the sketch of the proof at the beginning of this subsection, under the assumption that f and g are smoothly homotopic and y is a common regular value.

(c) Let y and z be two regular values for $\partial H : M \to N$. Since N is homogeneous (23.7), there is a diffeomorphism h from N to N that carries y to z. Furthermore, examining the proof of 23.7, we can see that the diffeomorphism h can be taken to be smoothly isotopic to the identity, see 26.24.

By 25.16, h preserves orientation, hence $\deg(\partial H, y) = \deg(h \circ (\partial H), h(y))$.

By (a), since $h \circ \partial H$ is smoothly homotopic to $\mathrm{id} \circ f$,

$$\deg(h \circ (\partial H), h(y)) = \deg(\mathrm{id} \circ (\partial H), h(y)) = \deg(\partial H, h(y)) = \deg(\partial H, z).$$

Thus $\deg(\partial H, y) = \deg(\partial H, z)$. Theorem 26.1 is proved.

The above exposition follows [Mil97]. There are more advanced treatments, such as in terms of transversal theory and intersection number in [GP74, p. 100, p. 107].

Applications

Proposition. *A disk cannot smoothly retract to its boundary.*

This is 24.5 for the case of D^{n+1}. Thus as in 24.7, it gives a proof for the smooth Brouwer fixed point theorem.

Proof. Suppose that there is such a retraction, a smooth map $f : D^{n+1} \to S^n$ that is the identity on S^n. Define $F : [0,1] \times S^n$ by $F(t,x) = f(tx)$. Then F is a smooth homotopy from a constant map to the identity map on the sphere. But these two maps have different degrees. □

Proposition (The fundamental theorem of Algebra). *Any non-constant polynomial with real coefficients has at least one complex root.*

Proof. Let $p(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_{n-1}z + a_n$, with $a_i \in \mathbb{R}$, $1 \le i \le n$. Suppose that p has no root, that is, $p(z) \neq 0$ for all $z \in \mathbb{C}$. As a consequence, $a_n \neq 0$.

For $t \in [0,1]$, let

$$q_t(z) = (1-t)^n z^n + a_1(1-t)^{n-1} t z^{n-1} + \cdots + a_{n-1}(1-t)t^{n-1}z + a_n t^n.$$

Observe that $q_t(z)$ is smooth with respect to the pair (t,z), and $q_0(z) = z^n$, $q_1(z) = a_n$.

Restricting z to the set $\{z \in \mathbb{C} \mid |z| = 1\} = S^1$, since $q_t(z) = t^n p((1-t)t^{-1}z)$ for $t \neq 0$, $q_t(z)$ has no roots. So $\frac{q_t(z)}{|q_t(z)|}$ is a smooth homotopy of maps from S^1 to itself, starting with the polynomial z^n and ending with the constant polynomial $\frac{a_n}{|a_n|}$. But these two maps have different degrees, see 26.9, a contradiction. □

26.7 Theorem (The Hairy Ball Theorem). *If n is even then every smooth tangent vector field on S^n has a zero.*

Example. On S^1, let $v : S^1 \to \mathbb{R}^2$, $v((x,y)) = (-y,x)$, then it is a tangent vector field on S^1 which is not zero anywhere.

In general, with odd $n = 2k+1$, similarly we can find a tangent vector field on S^n which is not zero anywhere, such as

$$v(x_1, x_2, x_3, x_4, \dots, x_{2k+2}) = (-x_2, x_1, -x_4, x_3, \dots, -x_{2k+2}, x_{2k+1}).$$

The above theorem says that this is not possible when the dimension of the sphere is even. Thus on the sphere S^2 any smooth tangent vector must vanish somewhere, explaining the name "Hairy Ball Theorem".

Proof. Suppose that v is a nonzero tangent smooth vector field on S^n. Let $w(x) = \frac{v(x)}{\|v(x)\|}$, then w is a unit smooth tangent vector field on S^n.

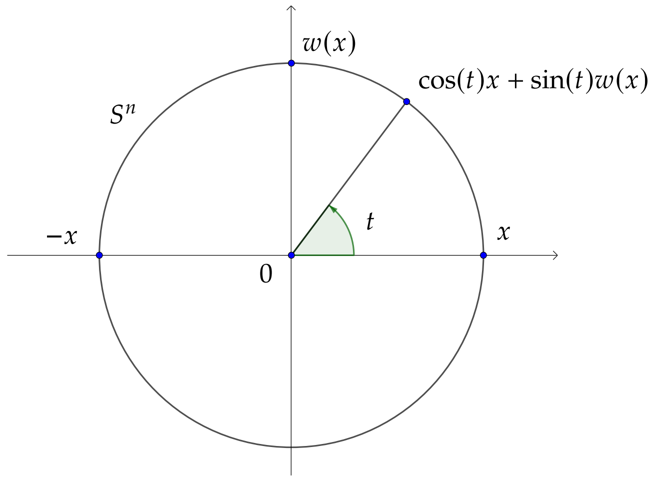


Figure 26.8: A homotopy on the sphere from x to $-x$.

Notice that $w(x)$ is perpendicular to x. On the plane spanned by x and $w(x)$ we can easily rotate vector x to vector $-x$, namely, let $F_t(x) = \cos(t)x + \sin(t)w(x)$ with $0 \leq t \leq \pi$, then F is a homotopy on S^n from x to $-x$. But the degrees of these two maps are different, see 26.17. $\square$

This notion of degrees of maps has other applications in the notion of indexes of vector fields and the Poincaré-Hopf index theorem [Mil97, p. 32], in homology [Hat01, section 2.2], in Nonlinear Analysis [Duc05, chương 8].

Problems

26.9. Let $f : S^1 \to S^1$, $f(z) = z^n$ with $n \in \mathbb{Z}$, using complex numbers, as in 21.17 and 26.2.

(a) Show that the derivative $df_{z=(x,y)}$ brings any tangent vector of S^1 to a tangent vector of S^1 of n times the length.

(b) Show that the derivative $df_{z=(x,y)}$ is orientation-preserving on the tangent space of S^1 at $z = (x, y)$ if and only if $n > 0$.

(c) Show that $\deg(f) = n$.

26.10. What happens if we drop the condition that N is connected in Theorem 26.1? Where do we use this condition?

26.11. Let M and N be oriented boundaryless manifolds, M is compact and N is connected. Let $f : M \to N$. Show that if $\deg(f) \neq 0$ then f is onto, i.e. the equation $f(x) = y$ always has a solution.

26.12. Let M be a compact, connected, oriented, boundaryless manifold. Let $f : M \to M$ be smooth.

(a) Show that if f is bijective then $\deg f = \pm 1$.

(b) Let $f^2 = f \circ f$. Show that $\deg(f^2) \geq 0$.

26.13. Compute the degree of

$$f : S^2 \to S^2$$
$$(x, y, z) \mapsto (y, x, z).$$

26.14. Compute the degree of

$$f : S^2 \to S^2$$
$$(x, y, z) \mapsto (y, -z, -x).$$

26.15. Let $r_i : S^n \to S^n$ be the reflection map

$$r_i((x_1, x_2, \dots, x_i, \dots, x_{n+1})) = (x_1, x_2, \dots, -x_i, \dots, x_{n+1}).$$

Compute $\deg(r_i)$.

26.16. Let $f : S^n \to S^n$ be the map that interchanges two coordinates:

$$f((x_1, x_2, \dots, x_i, \dots, x_j, \dots, x_{n+1})) = (x_1, x_2, \dots, x_j, \dots, x_i, \dots, x_{n+1}).$$

Compute $\deg(f)$.

26.17. Let $r : S^n \to S^n$ be the antipodal map

$$r((x_1, x_2, \dots, x_{n+1})) = (-x_1, -x_2, \dots, -x_{n+1}).$$

Check that $\deg(r) = (-1)^{n+1}$.

26.18. Let $f : S^4 \to S^4$, $f((x_1, x_2, x_3, x_4, x_5)) = (x_2, x_4, -x_1, x_5, -x_3)$. Find $\deg(f)$.

26.19. Construct a map from S^2 to itself of any given degree.

26.20. If $f, g : S^n \to S^n$ are smooth such that $f(x) \neq -g(x)$ for all $x \in S^n$ then f is smoothly homotopic to g.

26.21. Let $f : M \to S^n$ be smooth. Show that if $\dim(M) < n$ then f is homotopic to a constant map.

26.22 (Brouwer fixed point theorem for the sphere). Let $f : S^n \to S^n$ be smooth. Show that if $\deg(f) \neq (-1)^{n+1}$ then f has a fixed point.

26.23. Show that any map of from S^n to S^n of odd degree carries a certain pair of antipodal points to a pair of antipodal point.

26.24. Improving 23.7, deduce from the proof that on a connected smooth manifold there is a self diffeomorphism, smoothly isotopic to the identity, that brings any given point to any given point.

26.25. * Check that being smoothly homotopic is an equivalence relation.

27 Integration of real functions

Let M be a smooth m-manifold in $\mathbb{R}^k$ and let $f : M \to \mathbb{R}$ be smooth. The purpose of this section is to construct the integral $\int_M f$.

In the special case when M is an open subset of $\mathbb{R}^m$ it is clear that $\int_M f$ should be the usual Lebesgue integral of f. For the integral be a real number, not ∞, we may impose that either f has compact support or M is compact.

Now suppose that M is parametrized by a single parametrization $\varphi : \mathbb{R}^m \to M$. By the same arguments in Calculus of several variables when we construct integrals over curves and surfaces, what we need is the notion of the m-dimensional volume for parallelepiped spanned by the m vectors in $\mathbb{R}^k$, $\left(\frac{\partial \varphi}{\partial u_i}\right)_{1\le i\le m}$.

Volume of parallelepiped

We take this as the principle for the definition of m-dimensional measure in $\mathbb{R}^k$: If P is an m-dimensional vector subspace of $\mathbb{R}^k$ and $F = (f_1, f_2, \dots, f_m)$ is an orthonormal vector basis for P under the Euclidean inner product, then the m-dimensional measure in P written the basis F should be same as the m-dimensional measure in $\mathbb{R}^m$ written in the canonical basis $E = (e_1, e_2, \dots, e_m)$. More precisely, the m-dimensional measure of a set $D = \{a_1 f_1 + a_2 f_2 + \cdots + a_m f_m \mid (a_1, a_2, \dots, a_n) \in D' \subset \mathbb{R}^m\}$ should be equal to the m-dimensional Lebesgue measure of D'. In other words, if T is the linear isometry from P to $\mathbb{R}^m$ bringing F to E, then T preserves measure.

This principle implies, and is consistent with, the property that the m-dimensional Lebesgue measure is invariant under orthonormal transformations of $\mathbb{R}^m$.

Now we concentrate on the case of parallelepiped. A parallelepiped spanned by m vectors $(v_1, v_2, \dots, v_m)$ in P is the set $[v_1, v_2, \cdots, v_m] = \{\sum_{1\le i\le m} a_i v_i \mid a_i \in [0,1]\}$. The linear transformation T will bring this set to the subset of $\mathbb{R}^m$ given by $\{\sum_{1\le i\le m} a_i T(v_i) \mid a_i \in \mathbb{R}\}$. Thus now we should find the volume in the case of $\mathbb{R}^m$.

In $\mathbb{R}^m$, the change of variable formula for Lebesgue integration contains the important property that if $v_i = \sum_{j=1}^m v_{ji} e_j$ then

$$\mathrm{vol}[v_1, v_2, \cdots, v_m] = |\det(v_{ij})_{1\le i\le m, 1\le j\le m}|.$$

We have the following expression:

$$
\begin{aligned}
|\det(v_{ij})| &= \left(\det\left((v_{ij})^T \cdot (v_{ij})\right)\right)^{1/2} \\
&= \left(\det\left(\sum_{k=1}^{m} v_{ki}v_{kj}\right)\right)^{1/2} \\
&= \left(\det\left(v_i \cdot v_j\right)_{ij}\right)^{1/2}.
\end{aligned}
$$

Example. When $m = 2$ we have had this formula derived by an elementary method:

$$
\begin{aligned}
|a \times b| &= |a||b| \sin \widehat{(a,b)} = \sqrt{|a|^2|b|^2(1-\cos^2 \widehat{(a,b)})} = \sqrt{|a|^2|b|^2 - \langle a,b\rangle^2} \\
&= \left[\det\begin{pmatrix} a \cdot a & a \cdot b \\ b \cdot a & b \cdot b \end{pmatrix}\right]^{1/2}.
\end{aligned}
$$

Returning to the case of m vectors in $\mathbb{R}^k$, now we have $\mathrm{vol}[v_1, v_2, \cdots, v_m] = \left(\det\left(T(v_i) \cdot T(v_j)\right)_{ij}\right)^{1/2}$. Notice that T preserves inner product, finally we obtain an answer:

$$
\mathrm{vol}[v_1, v_2, \cdots, v_m] = \left(\det\left(v_i \cdot v_j\right)_{ij}\right)^{1/2}.
$$

Integration of real functions for single parametrizations

Now it is clear what we should do:

Definition. *Suppose that $M \subset \mathbb{R}^k$ is parametrized by a single parametrization $\varphi : \mathbb{R}^m \to M$. Let $f : M \to \mathbb{R}$. Let $g_{ij} = \frac{\partial\varphi}{\partial u_i} \cdot \frac{\partial\varphi}{\partial u_j}$. We define*

$$
\int_M f = \int_{\mathbb{R}^m} f \circ \varphi \sqrt{\det(g_{ij})_{ij}}.
$$

To guarantee that this integral exists as a real number we may require f to have compact support.

Example. When $m = 1$ this is path integral, and when $m = 2$ this is surface integral, as studied in Calculus.

Lemma. *This definition is independence from the choice of parametrization.*

Proof. Let $\psi : \mathbb{R}^m \to M$ be another parametrization. Applying the change of variable $\varphi^{-1} \circ \psi$ and using the change of variable formula in Lebesgue integration:

$$\begin{aligned}
\int_{\mathbb{R}^m} f\circ\varphi\sqrt{\det(g_{ij})} &= \int_{\mathbb{R}^m} f\circ\varphi\sqrt{\det\left(J_\varphi^T J_\varphi\right)} \\
&= \int_{\mathbb{R}^m}\left(f\circ\varphi\sqrt{\det\left({J_\varphi}^T J_\varphi\right)}\right)\circ(\varphi^{-1}\circ\psi)|\det J_{\varphi^{-1}\circ\psi}| \\
&= \int_{\mathbb{R}^m}(f\circ\varphi)\circ(\varphi^{-1}\circ\psi)\sqrt{\det\left({J_{\varphi^{-1}\circ\psi}}^T\left(({J_\varphi}^T J_\varphi)\circ(\varphi^{-1}\circ\psi)\right)J_{\varphi^{-1}\circ\psi}\right)} \\
&= \int_{\mathbb{R}^m}(f\circ\psi)\sqrt{\det\left(\left(J_\varphi\circ(\varphi^{-1}\circ\psi)J_{\varphi^{-1}\circ\psi}\right)^T\left(J_\varphi\circ(\varphi^{-1}\circ\psi)J_{\varphi^{-1}\circ\psi}\right)\right)} \\
&= \int_{\mathbb{R}^m}(f\circ\psi)\sqrt{\det\left({J_\psi}^T J_\psi\right)}.
\end{aligned}$$

□

Remark. In the language of measure theory, we may say that the measurable sets on M consists of the sets $\varphi(V)$ where V is Lebesgue measurable on $\mathbb{R}^m$, and the measure of such a set is $\mathrm{vol}(\varphi(V)) = \int_V \sqrt{\det(g_{ij})}d\mu$ where μ is the Lebesgue measure of $\mathbb{R}^m$.

Partition of unity

When it is not possible to cover the manifold by a single parametrization, we will use partitions of unity.

27.1 Theorem (smooth partition of unity). *Let M be a compact smooth manifold. Let O be any open cover of M. There exists a family of functions $(f_i)_{1\le i\le k}$ such that $f_i : M \to \mathbb{R}$ is smooth, $0 \le f_i \le 1$, $\mathrm{supp}(f_i) = \overline{\{x \in M \mid f_i(x) \ne 0\}} \subset U$ for some $U \in O$, and $\sum_{1\le i\le k} f_i = 1$.*

Proof. The proof follows the steps below:

(a) For each $x \in M$, show that there are open sets V_x and W_x in M such that $x \in V_x \subset \overline{V}_x \subset W_x \subset \overline{W}_x \subset U$ for some $U \in O$.

(b) The collection $(V_x)_{x\in M}$ is an open cover of M with a finite subcover $(V_i)_{1\le i\le k}$. There is a smooth function $\varphi_i : M \to \mathbb{R}$ such that $0 \le \varphi_i \le 1$, $\varphi_i\big|_{V_i} \equiv 1$, $\mathrm{supp}(\varphi_i) \subset \overline{W}_i \subset U$ for some $U \in O$.

(c) Let $f_i = \frac{\varphi_i}{\sum_{i=1}^k \varphi_i}$. Then $(f_i)_{1\le i\le k}$ is the family of functions we are looking for.

□

Integration of real functions: general cases

Let $(p_i)_{1\le i\le k}$ be any finite partition of unity associated to any cover of M by coordinate neighborhoods. Since $\sum_{i=1}^k p_i = 1$, we can write $f = \left(\sum_{i=1}^k p_i\right) f = \sum_{i=1}^k p_i f$. Since integral should be linear, we can propose:

Definition. *Let M be a compact manifold. Let $(p_i)_{1\le i\le n}$ be any finite partition of unity associated to any cover of M by coordinate neighborhoods. We define*

$$\int_M f = \sum_{i=1}^n \int_M p_i f.$$

Lemma. *The definition of integral is independence from the choice of partition of unity.*

Proof. Let $(q_j)_{1\le j\le l}$ be any partition of unity associated to any cover of M by coordinate neighborhoods. We have

$$\sum_i \int_M p_i f = \sum_i \int_M (\sum_j q_j)(p_i f) = \sum_i \int_M \sum_j p_i q_j f = \sum_{ij} \int_M p_i q_j f.$$

Similarly

$$\sum_j \int_M p_j f = \sum_{ij} \int_M p_i q_j f.$$

Thus $\sum_i \int_M p_i f = \sum_j \int_M p_j f$. ∎

As a special case, we can define the **volume** of a compact manifold $M \subset \mathbb{R}^k$ to be $\int_M 1$.

Integration of real functions is a special case of a far-reaching theory of integration of differential forms on manifolds. In this theory important results of differential and integral calculus such as the Newton–Leibniz formula and the Stokes formula are generalized to higher dimensions. The reader can study this theory from sources such as [Spi65], [GP74], [VSt].

Problems

27.2 (generalized cross-product). In $\mathbb{R}^3$ we have the cross-product of two vectors $v_1 \times v_2$. Here we attempt to generalize it to a cross-product of $(n-1)$ vectors in $\mathbb{R}^n$ with similar properties [Spi65, p. 83]. Let $v_1, v_2, \dots, v_{n-1}$ be vectors in $\mathbb{R}^n$, we want to define a vector $v = v_1 \times v_2 \times \cdots \times v_{n-1} \in \mathbb{R}^n$ such that:

(a) For each $1 \le i \le n-1$, $v \perp v_i$,

(b) The orientation of v is such that if $v_1, v_2, \dots, v_{n-1}$ are linearly independent then $(v_1, v_2, \dots, v_{n-1}, v)$ is a positive basis of $\mathbb{R}^n$,

(c) The length of v is equal to the volume of the $(n-1)$-dimensional parallelepiped formed by the vectors $v_1, v_2, \ldots, v_{n-1}$.

(d) The map $(v_1, v_2, \ldots, v_{n-1}) \mapsto v = v_1 \times v_2 \times \cdots \times v_{n-1}$ is linear in each variable (multilinear).

Other topics

Many topics for course projects and further study are already suggested in previous sections. Below there are several more topics.

- Classification of one dimensional manifolds, [Mil97, p. 55], [GP74, p. 208].
- Sketch of a proof of Sard theorem, [Mil97, p. 16].
- Modulo 2 degrees of maps, without using orientation, [Mil97, p. 20].
- Indexes of vector fields, [Mil97, p. 32], [GP74, p. 132].
- From the finite dimensional Brouwer fixed point theorem to the infinite dimensional Schauder fixed point theorem, [D. Gilbarg, N. Trudinger, *Elliptic partial differential equations of second orders*, Springer, 2001, p. 279].
- Collect and compare some proofs of the Fundamental Theorem of Algebra.
- The notion of transversal intersection of smooth manifolds, [GP74, p. 27].
- The notions of immersion, submersion, embedding of smooth manifolds, [GP74, p. 12, p. 20].
- Approximation of continuous maps by smooth maps, [Lee13, p. 136].
- For an overview of developments of topology, there is a video record of John Milnor's lecture in 2014, *Topology through Four Centuries*, at `https://www.youtube.com/watch?v=oMpCOJbaoaY`

Suggestions for some problems

1.12 Let A be infinite, and let A_n be the set of subsets of A with n elements. $A_n \neq \emptyset$ by induction. Choose a sequence $(S_n)_{n\in\mathbb{Z}^+}$ such that $S_n \in A_n$, by the Axiom of choice. Let $S = \bigcup_{n\in\mathbb{Z}^+} S_n$.

1.13 Use 1.12.

1.20 $\bigcup_{n=1}^{\infty}[n, n+1] = [1, \infty)$.

1.21 Use the idea of the Cantor diagonal argument in the proof of 1.5. In this case the issue of different presentations of same real numbers does not appear.

1.22 Use the injective map $g \circ f$.

1.25 For $A \subset \mathbb{Z}^+$, if $n \in A$ let $a_n = 1$, otherwise let $a_n = 0$. Consider the map $A \mapsto a = a_1 a_2 \dots a_n \dots$. Use 1.23 and 1.22.

1.26 Proof by contradiction.

2.25 Show that each ball in one metric contains a ball in the other metric with the same center.

2.27 (a) Notice that d_1 is bounded. (b) Check that both the function $\frac{x}{1+x}$ and its inverse function are increasing for $x \geq 0$. Check that $d_2 \geq \frac{1}{2}d_1$.

3.30 Use 3.29 to construct a homeomorphism bringing $\{0\} \times [0, 1]$ to $\{0\} \times [\frac{1}{2}, 1]$, and $[0, 1] \times \{0\}$ to $\{0\} \times [0, \frac{1}{2}]$.

3.33 See 2.26 and 3.3.

3.34 Let $f : \partial D^n \to \partial D^n$ be a homeomorphism. Consider $F : D^n \to D^n$, $F(x) = \|x\| f(\frac{x}{\|x\|})$ (radial extension).

3.28 Compare the sub-interval $[1, 2\pi)$ and its image via φ.

3.32 For two triangles ABC and $A'B'C'$, and for $\alpha, \beta, \gamma \in [0, 1]$ such that $\alpha + \beta + \gamma = 1$, consider $f(\alpha A + \beta B + \gamma C) = (\alpha A' + \beta B' + \gamma C')$. Check that if $P = \alpha A + \beta B + \gamma C$ then α, β, γ depend continuously on P. Write $\overrightarrow{AP} = \beta\overrightarrow{AB} + \gamma\overrightarrow{AC}$, then β and γ are solutions to a system of linear equations, hence depend continuously on the coefficients.

3.38 See 2.26

4.25 Use the characterization of connected subspaces of the Euclidean line.

4.14 Let A be countable and $x \in \mathbb{R}^2 \setminus A$. There is a line passing through x that does not intersect A (by an argument involving countability of sets).

4.32 Use 3.29 to modify each letter part by part. See 3.30. Use connectedness to distinguish spaces.

5.1 Let C be a countable subset of $[0, \Omega)$. The set $\bigcup_{c \in C}[0, c)$ is countable while the set $[0, \Omega)$ is uncountable. This implies C is bounded from above.

5.27 Consider the set of all irrational numbers.

6.12 This is a special case of 6.5.

6.13 Use Lebesgue number.

6.3 See the proof of 6.4.

6.18 Use 6.3.

6.19 Use 6.18.

6.20 Let X be a compact metric space, and let I be an open cover of X. For each $x \in X$ there is an open set $U_x \in I$ containing x. There is a number $\epsilon_x > 0$ such that the ball $B(x, 2\epsilon_x)$ is contained in U_x. The collection $\{B(x, \epsilon_x) \mid x \in X\}$ is an open cover of X, therefore there is a finite subcover $\{B(x_i, \epsilon_i) \mid 1 \leq i \leq n\}$. Let $\epsilon = \min\{\epsilon_i \mid 1 \leq i \leq n\}$.

6.21 (b) $B(x, r - d(x, y)) \subset B(y, r)$.

6.24 Use 6.23.

6.31 Suppose $x \in U$. Let $x \in V$ and $\bar{V}$ be compact. Using 6.3, there are W_1 and W_2 separating x and $\bar{V} \setminus U$. Let $W = W_1 \cap U \cap V$. Check that $\overline{W} \subset U$.

6.32 Use 6.31.

6.34 Use Alexandroff compactification.

6.35 Show that if $Y = \bigcap_{i=1}^{\infty} X_i$ is not connected then there are two disjoint sets U and V which are open in X such that $Y \subset U \cup V$, $U \cap Y \neq \emptyset$, and $V \cap Y \neq \emptyset$, using 6.19. Show that $U \cup V$ contains X_n for some n, by considering the sequence $(X_n \setminus (U \cup V))_n$ and using 6.12.

7.8 Look at their bases.

7.13 For open sets, consider the cases of finite products and infinite products separately. For closed sets, use nets.

7.14 For open sets, show that the projection of an element of the basis is open.

7.18 Use 7.5 to prove that the inclusion map is continuous.

7.19 Use 7.18.

7.20 Let (x_i) and (y_i) be in $\prod_{i \in I} X_i$. Let γ_i be a continuous path from x_i to y_i. Let $\gamma = (\gamma_i)$.

7.21 (b) Use 7.18. (c) Fix a point $x \in \prod_{i \in I} X_i$. Use (b) to show that the set A_x of points that differs from x at at most finitely many coordinates is connected. Furthermore A_x is dense in $\prod_{i \in I} X_i$.

7.22 Use 7.18. It is enough to prove for the case an open cover of $X \times Y$ by open sets of the form a product of an open set in X with an open set in Y. For each "slice" $\{x\} \times Y$ there is a finite subcover $\{U_{x,i} \times V_{x,i} \mid 1 \leq i \leq n_x\}$. Take $U_x = \bigcap_{i=1}^{n_x} U_{x,i}$. The collection $\{U_x \mid x \in X\}$ covers X so there is a subcover $\{U_{x_j} \mid 1 \leq j \leq n\}$. The collection $\{U_{x_j,i} \times V_{x_j,i} \mid 1 \leq i \leq n_{x_j},\ 1 \leq j \leq n\}$ is a finite subcover of $X \times Y$.

7.28 (b) Use 5.4. (c) Condider the corresponding bases. Let $U = \prod_{n \in \mathbb{Z}^+} U_n$ where $U_n = B(x_n, \epsilon_n)$ such that $U_n = [0,1]$ for $n \notin \{n_k \mid 1 \leq k \leq N\}$. Check that if $\frac{1}{2^N} < \frac{\epsilon}{2}$ then $B_d(x, \epsilon) \supset \left(\prod_{n=1}^{N} B\left(x_n, \frac{\epsilon}{2}\right)\right) \times \prod_{n=N+1}^{\infty} [0,1]$. Check that if $\epsilon = \min\{\epsilon_{n_k} \mid 1 \leq k \leq N\}$ then $U \supset B_d\left(x, \frac{\epsilon}{2^{n_N}}\right)$.

9.23 ($\Leftarrow$) Use 6.19 and the Urysohn lemma 9.7.

9.12 Use 6.32 and 6.17 and the proof of Urysohn lemma.

9.20 Use 9.19.

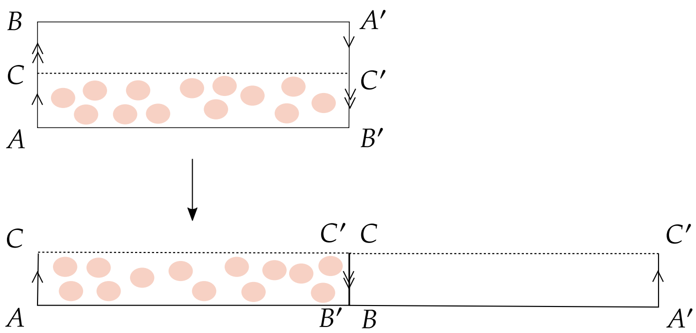
9.4 See 7.4 and 7.6.

9.16 Use 6.31. Use a technique similar to the one in 7.22.

8.32 Cut the square by a suitable diagonal, then glue back the resulting two triangles at a different pair of edges.

8.34 Cut one of the two squares by a suitable diagonal, then glue a different pair of edges of the resulting two triangles.

8.36 To give a rigorous argument we can simply describe the figure below.



The map from $X = ([0,1] \times [0,1]) \setminus \left([0,1] \times \{\frac{1}{2}\}\right)$ to $Y = [0,2] \times [0, \frac{1}{2})$ given by

$$(x, y) \mapsto \begin{cases} (x, y), & y < \frac{1}{2}, \\ (x+1, 1-y), & y > \frac{1}{2}, \end{cases}$$

is bijective and is continuous. The induced map to $Y/(0,y) \sim (2,y)$ is surjective and is continuous. Then its induced map on $X/(0,y) \sim (1,1-y)$ is bijective and is continuous, hence is a homeomorphism between $X/(0,y) \sim (1,1-y)$ and $Y/(0,y) \sim (2,y)$.

8.47 The idea is easy to be visualized in the cases $n = 1$ and $n = 2$. Let $S^+ = \{x = (x_1, x_2, \dots, x_{n+1}) \in S^n \mid x_1 \geq 0\}$, the upper hemisphere. Let $S^0 = \{x = (x_1, x_2, \dots, x_{n+1}) \in S^n \mid x_1 = 0\}$, the equator. Let $f : S^n \to S^+$ be given by $f(x) = x$ if $x \in S^+$ and $f(x) = -x$ otherwise. Then the following diagram is commutative:

$$\begin{tikzcd} S^n \arrow[r, "f"] \arrow[d] \arrow[dr, "p\circ f"] & S^+ \arrow[d, "p"] \\ S^n/x \sim -x \arrow[r, "\tilde{f}"] & S^+/x \sim -x,\ x \in S^0 \end{tikzcd}$$

Then it is not difficult to show that $S^+/x \sim -x,\ x \in S^0$ is homeomorphic to $\mathbb{R}\mathrm{P}^n = D^n/x \sim -x,\ x \in \partial D^n$.

8.52 Let h be the map induced by f, then for an open set $O \subset X/\sim$

$$\begin{aligned} h(O) &= \{h([x]) \mid [x] \in O\} = \{f(x) \mid [x] \in O\} \\ &= \{f(x) \mid x \in p^{-1}(O)\} = f(p^{-1}(O)) \end{aligned}$$

therefore $h(O)$ is open and so h is an open map.

8.53 Use this diagram, where i is the inclusion map and p_1 and p_2 are quotient maps, and use 8.3 to check that h is a homeomorphism:

$$\begin{tikzcd} {[0,1]} \arrow[r, "i"] \arrow[d, "p_1"] \arrow[dr, "p_2\circ i"] & \mathbb{R} \arrow[d, "p_2"] \\ {[0,1]_{/0\sim 1}} \arrow[r, "h"] & \mathbb{R}/_{x\sim x+n,\ n\in\mathbb{Z}} \end{tikzcd}$$

Another approach is to use 8.52. To check that $f : \mathbb{R} \to S^1, f(t) = (\cos n2\pi t, \sin n2\pi t)$ is an open map, it is sufficient to check that if the interval (a,b) is small enough then $f(a,b)$ is an open subset of S^1.

8.54 Similar to 8.53.

8.57 There is at least one equivalence relation containing R, and take the intersection of all equivalence relations containing R.

8.59 Examine the following commutative diagrams, and use 8.2 to check that the maps are continuous. Here $g([x]_{R_1}) = [x]_R$.

$$\begin{tikzcd} X \arrow[r, "p_1"] \arrow[d, "p"] \arrow[dr, "f=p_2\circ p_1"] & X/R_1 \arrow[d, "p_2"] \\ X/R \arrow[r, "\tilde{f}"] & (X/R_1)/\tilde{R}_2 \end{tikzcd}$$

$$\begin{tikzcd}
X \arrow[r, "p_1"] \arrow[d, "p"] & X/R_1 \arrow[d, "p_2"] \arrow[dl, "g"] \\
X/R & (X/R_1)/\tilde{R}_2 \arrow[l, "\tilde{g}"]
\end{tikzcd}$$

10.19 One approach is by using ideas in 3.33. A generalization is in 17.22.

11.22 Deleting an open disk is the same as deleting the interior of a triangle.

11.25 Count edges from of the set of triangles. Count vertices from the set of triangles. Count edges from the set of vertices, notice that each vertex belongs to at most $(v-1)$ edges.

14.9 Use the fact that the torus T^2 is homeomorphic to $S^1 \times S^1$ (8.25).

15.7 Take a rectangle containing the set of points to be deleted. The plane has a deformation retraction to that rectangle. Divide the rectangle into finitely many sub-rectangles such that each sub-rectangle contains at most one point to be deleted, which is in the interior. A rectangle with an interior point deleted has a deformation retraction to its boundary. The new space has a deformation retraction to a bouquet of circles, for example by first taking a homotopy collapsing the all the vertical edges of the sub-rectangles.

15.10 Consider mapping the group $\langle a, b \mid a^2b^2 = 1\rangle$ to the dihedral group D_3, the group of isometries of an equilateral triangle, which is also the group S_3 of permutations of three elements. In particular consider mapping a and b to any two elements of D_3 which are reflections about a bisector of an angle, i.e. transposition elements of S_3, such as $(1,2,3) \mapsto (1,3,2)$ and $(1,2,3) \mapsto (3,2,1)$. For more, see [Fra14, p. 79].

16.6 Let X be the simplicial complex, A be the set of vertices path-connected to the vertex v_0 by edges, U be the union of simplices containing vertices in A. Let B be the set of vertices not path-connected to the vertex v_0 by edges, V be the union of simplices containing vertices in B. Then U and V are closed and disjoint.

16.7 Use problem 16.6.

17.11 First take a deformation retraction to a sphere.

17.12 Show that $\mathbb{R}^3 \setminus S^1$ is homotopic to Y which is a closed ball minus a circle inside. Show that $Y = S^1 \vee S^2$, see [Hat01, p. 46]. Or write Y as a union of two halves, each of which is a closed ball minus a straight line, and use the Van Kampen theorem.

17.16 Use Mayer-Vietoris sequence.

17.22 Let S be a convex compact subset of $\mathbb{R}^n$. Suppose that $p_0 \in \operatorname{Int} S$. For $x \in S^{n-1}$, let $f(x)$ be the intersection of the ray from p_0 in the direction of x with the boundary of S, namely let $t_x = \max\{t \in \mathbb{R}^+ \mid p_0 + tx \in S\}$ and let $f(x) = p_0 + t_x x$. Notice that the straight segment $[p_0, f(x)]$ has only $f(x)$ as a boundary point of S, since if we take $B(p_0, \epsilon) \subset S$ then the convex hull of the set $\{f(x)\} \cup B(p_0, \epsilon)$ contains

$[p_0, f(x))$ in the interior. Check that the map $g : D^n \to S$ sending the straight segment $[0, x]$ linearly to the straight segment $[p_0, f(x)]$ is well-defined and is bijective. To check that g is a homeomorphism it might be easier to consider the continuity of the inverse map. Also $f : S^{n-1} \to \partial S$ is a homeomorphism.

19.11 Consider a neighborhood of $(0, 0)$. Any continuous path from a point (x_1, y_1) to a point (x_2, y_2), where $x_1 < 0, x_2 > 0$ must pass through $(0, 0)$.

19.17 See 4.8.

20.9 For the existence of q, it is sufficient to consider the distance function restricted to a compact set, such as $B'(p, r) \cap M$ with sufficient large r. Consider any smooth path passing through q. Differentiate the square of the distance function.

21.18 Use 6.9.

21.17 Use the Cauchy–Riemann equations for derivatives of complex functions.

21.19 Each $x \in f^{-1}(y)$ has a neighborhood U_x on which f is a diffeomorphism. Let $V = [\bigcap_{x \in f^{-1}(y)} f(U_x)] \setminus f(M \setminus \bigcup_{x \in f^{-1}(y)} U_x)$. Consider $V \cap S$.

21.23 Consider

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ \varphi\uparrow & & \uparrow\psi \\ U & \xrightarrow{\psi^{-1}\circ f\circ\varphi} & V \end{array}$$

Since df_x is surjective, it is bijective. Then $d(\psi^{-1} \circ f \circ \varphi)_u = d\psi^{-1}_{f(x)} \circ df_x \circ d\varphi_u$ is an isomorphism. The Inverse function theorem can be applied to $\psi^{-1} \circ f \circ \varphi$.

21.24 We need to prove that φ^{-1} is smooth. Let ψ be a local parametrization of a neighborhood of $x = \varphi(u) \in M$. The set $U = \varphi^{-1}(\varphi(\mathbb{R}^m) \cap \psi(\mathbb{R}^m))$ is an open neighborhood of u in $\mathbb{R}^m$. The function $\psi^{-1} \circ \varphi|_U$ is a diffeomorphism from U to its image W. This implies $\varphi|_U^{-1} = (\psi^{-1} \circ \varphi|_U)^{-1} \circ \psi^{-1}$ is smooth on $\psi(W)$, an open neighborhood of x in M.

21.25 Use the Implicit function theorem.

21.28 Use Problem 21.22.

22.1 Use 21.22, or the following direct argument. Let α be a smooth path in $f^{-1}(c)$ through $x = \alpha(0)$ then for all t we have $f(\alpha(t)) = a$, implying $\nabla f(\alpha(t)) \cdot \alpha'(t) = 0$, thus $\nabla f(\alpha(t)) \perp \alpha'(t)$, in particular $\nabla f(x) \perp \alpha'(0)$. Since $\alpha'(0)$ is an arbitrary tangent vector of $f^{-1}(c)$ we have $\nabla f(x) \perp T_x f^{-1}(c)$.

23.9 Show that $f_1^{(n)}(x) = e^{-1/x} P_n(1/x)$ for $x > 0$, where P_n is a polynomial.

24.14 Consider $\mathbb{R}^{m-1} \hookrightarrow \partial\mathbb{H}^m \xrightarrow{f} \mathbb{R}^n$.

24.15 Use Sard theorem.

24.18 Use 23.5. Show that the flow ϕ is injective, and is surjective. Notice that $f(\phi_t(x)) = t + a$ for $x \in f^{-1}(a)$ and $0 \le t \le b - a$.

25.15 We need that for every list of independent vectors $v_1, v_2, \dots, v_{n-1}$ in $\mathbb{R}^n$ there is a unique non-zero vector $v_n \in \mathbb{R}^n$ that is perpendicular to and depends smoothly on $v_1, v_2, \dots, v_{n-1}$. The vector v_n could be constructed using the Gram–Schmidt orthonormalization procedure. Another approach is to take the generalized cross product as in 27.2.

25.16 Consider the case of maps from $\mathbb{R}^m$ to $\mathbb{R}^m$ first, here $\frac{\partial F}{\partial x}$ is smooth with respect to (x,t), hence $\det \frac{\partial F}{\partial x}(x,t) \neq 0$ is continuous with respect to t, thus its sign does not change with t, in particular at $t = 0$ and $t = 1$. For maps on manifolds, use parametrizations. Let φ be an orientation-preserving parametrization of a neighborhood of $x \in M$. Using Lebesgue number, there are a sequence of neighborhoods $(V_i)_{1\le i\le k}$ parametrized by orientation-preserving ψ_i and a partition $0 = t_0 < t_1 < \cdots < t_k = 1$ such that $F_t(x) \in V_i$ for $t_i \le t \le t_{i+1}$. Repeat the previous argument with $\psi_i^{-1}(F(\varphi(u), t)$ instead of F.

26.19 Define a map on the sphere S^2 such that on each level (latitude) of the sphere S^2 it is given by a map similar to the map in 26.9.

26.20 Note that $f(x)$ and $g(x)$ will not be antipodal points. Use the homotopy

$$F_t(x) = \frac{(1-t)f(x) + tg(x)}{\left\|(1-t)f(x) + tg(x)\right\|}.$$

26.21 Using Sard Theorem, show that f cannot be onto.

26.22 If f does not have a fixed point then f will be homotopic to the reflection map.

26.24 Modify the proof of 23.7. Locally notice that the homotopy F is actually a smooth isotopy. Globally, repeat the argument with diffeomorphism isotopic to the identity instead of just diffeomorphism.

26.25 We need to make sure that the homotopies being used are smooth. To check that being smoothly homotopic is transitive, to glue smoothly two smooth homotopies F_t and G_t, take

$$H_t = \begin{cases} F_{\varphi(2t)}, & 0 \le t \le \frac{1}{2}, \\ G_{\varphi(2t-1)}, & \frac{1}{2} \le t \le 1, \end{cases}$$

where $\varphi : [0,1] \to [0,1]$ is smooth such that it is 0 on an interval $[0,b)$ and 1 on an interval $(a,1]$, see Problem 23.9.

27.2 Let $v_1 \times v_2 \times \cdots \times v_{n-1} \in \mathbb{R}^n$ be defined by the formula

$$\forall v \in \mathbb{R}^n, (v_1 \times v_2 \times \cdots \times v_{n-1}) \cdot v = \det(v_1, v_2, \dots, v_{n-1}, v)$$

using a finite dimension case of the Riesz representation theorem for linear functional. Another approach is to generalize the symbolic determinant formula for the cross product in $\mathbb{R}^3$.

Bibliography

[AF08] Colin Adams and David Fransoza, *Introduction to topology: Pure and applied*, Prentice Hall, 2008.

[Ams83] M. A. Amstrong, *Basic topology*, Springer, 1983.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, American Mathematical Society, 2001.

[Bre93] Glen E. Bredon, *Topology and Geometry*, Graduate Texts in Mathematics, vol. 139, Springer, 1993.

[Bre11] Haim Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Springer, 2011

[Cai94] George L. Cain, *Introduction to general topology*, Addison-Wesley, 1994.

[Con90] John B. Conway, *A course in functional analysis*, Springer-Verlag, 1990.

[Cro78] Fred H. Croom, *Basic concepts of algebraic topology*, Springer-Verlag, 1978.

[dC76] Manfredo do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, 1976.

[DFN85] Boris A. Dubrovin, Anatoly T. Fomenko, and Sergey P. Novikov, *Modern Geometry–Methods and Applications, Part II*, Springer-Verlag, 1985.

[DFN90] Boris A. Dubrovin, Anatoly T. Fomenko, and Sergey P. Novikov, *Modern Geometry–Methods and Applications, Part III*, Springer-Verlag, 1990.

[Dun18] Bjorn Dundas, *A short course on differential topology*, Cambridge University Press, 2018.

[Duc05] Dương Minh Đức, *Giải tích hàm*, Nhà xuất bản Đại học Quốc gia Thành phố Hồ Chí Minh, 2005.

[Dug66] James Dugundji, *Topology*, Allyn and Bacon, 1966.

[EH10] Herbert Edelsbrunner and John L Harer, *Computational topology: an introduction*, AMS, 2010.

[End77] Herbert B. Enderton, *Elements of set theory*, Academic Press, 1977.

[Eng89] Ryszard Engelking, *General topology*, Heldermann Verlag, 1989.

[Fra14] John B. Fraleigh, *A first course in abstract algebra* , 7th edition, Pearson 2014.

[Gal10] Joseph A. Gallian, *Contemporary abstract algebra*, Brooks/Cole, 2010.

[GT01] Jonathan L. Gross and Thomas W. Tucker, *Topological Graph Theory*, Dover, 2001.

[GP74] Victor Guillemin and Alan Pollack, *Differential topology*, Prentice-Hall, 1974.

[Hal74] Paul R. Halmos, *Naive Set Theory*, Springer, 1974.

[Hat01] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, 2001, `https://pi.math.cornell.edu/~hatcher/`

[Hir76] Morris W. Hirsch, *Differential Topology*, Springer, 1976.

[HS74] Morris W. Hirsch and Stephen Smale, *Differential equations, dynamical systems, and linear algebra*, Academic Press, 1974.

[HJ99] Karen Hrbacek and Thomas Jech, *Introduction to set theory*, 3rd ed., Marcel Dekker, 1999.

[Hun74] Thomas W. Hungerford, *Algebra*, Holt, Rinehart and Winston, Inc., 1974.

[HY61] John G. Hocking and Gail S. Young, *Topology*, Dover, 1961.

[J84] Klaus Jänich, *Topology*, Undergraduate Texts in Mathematics, Springer-Verlag, 1984.

[Kel55] John L. Kelley, *General topology*, Van Nostrand, 1955.

[KF75] A. N. Kolmogorov and S. V. Fomin, *Introductory real analysis*, Dover Publications Inc., 1975.

[KMM04] Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek, *Computational homology*, Springer, 2004.

[Kui20] Alexander Kupers, *Lectures on differential topology*, `https://www.utsc.utoronto.ca/people/kupers`

[Lan87] Serge Lang, *Linear Algebra*, 3rd ed., Springer, 1987.

[Lan97] Serge Lang, *Undergraduate analysis*, 2nd ed., Springer, 1997.

[Lee11] John M. Lee, *Introduction to topological manifolds*, 2nd ed., Springer, 2011.

[Lee13] John M. Lee, *Introduction to smooth manifolds*, 2nd ed., Springer, 2013.

[Mas91] William S. Massey, *A basic course in algebraic topology*, Springer, 1991.

[Mil97] John Milnor, *Topology from the differentiable viewpoint*, Princeton University Press, 1997.

[Moi77] Edwin E. Moise, *Geometric topology in dimensions 2 and 3*, Springer-Verlag, 1977.

[MT01] Bojan Mohar and Carsten Thomassen, *Graphs on surfaces*, JHU Press, 2001.

[Mor24] Sydney A. Morris, *Topology without tears*, `https://www.topologywithouttears.net`

[Mun84] James Munkres, *Elements of algebraic topology*, Addison-Wesley, 1984.

[Mun91] James Munkres, *Analysis on manifolds*, Addison-Wesley, 1991.

[Mun00] James Munkres, *Topology a first course*, 2nd ed., Prentice-Hall, 2000.

[Qui] Gereon Quick, *Differential Topology Lecture Notes*, `https://folk.ntnu.no/gereonq/DiffTop_Lecture_Notes.pdf`

[Ros99] Dennis Roseman, *Elementary topology*, Prentice-Hall, 1999.

[Rud86] Walter Rudin, *Real and complex analysis*, 3rd ed., McGraw Hill, 1986.

[Sas11] Anant R. Sastri, *Elements of Differential Topology*, CRC Press, 2011.

[SJ70] Lynn A. Steen and J. Arthur Seebach Jr., *Counterexamples in topology*, Holt, Rinehart and Winston, 1970.

[Spi65] Michael Spivak, *Calculus on manifolds*, Addison-Wesley, 1965.

[Spi99] Michael Spivak, *A comprehensive introduction to differential geometry*, vol. 2, Publish or Perish, 1999.

[TTV] Đinh Ngọc Thanh, Bùi Lê Trọng Thanh, Huỳnh Quang Vũ, *Bài giảng Giải tích hàm*, `https://sites.google.com/view/hqvu/teaching`

[Tu13] Loring W. Tu, *Introduction to manifolds*, Springer, 2nd ed., 2013.

[Vas01] V. A. Vassiliev, *Introduction to Topology*, American Mathematical Society, 2001.

[Vic94] James W. Vick, *Homology theory: an introduction to algebraic topology*, 2nd ed., Springer-Verlag, 1994.

[Vgt3] Huỳnh Quang Vũ, *Bài giảng Tích phân bội và Giải tích vectơ*, `https://sites.google.com/view/hqvu/teaching`

[VSt] Huỳnh Quang Vũ, *Bài giảng Dạng vi phân và Công thức Stokes*, `https://sites.google.com/view/hqvu/teaching`

[Vhhvp] Huỳnh Quang Vũ, *Bài giảng Hình học vi phân của đường và mặt*, `https://sites.google.com/view/hqvu/teaching`

[Wee02] J. Weeks, *The shape of space*, Marcel Dekker, 2002.

Index