Math in Moscow

A. B. Sossinsky

TOPOLOGY-I:
Lecture Notes

Independent University
of Moscow • 20??

Alexey Bronislavovich Sossinsky.

Topology-I: Lecture Notes.

????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ?????

* * *

Алексей Брониславович Сосинский.

Топология-I: записи лекций
(на английском языке).

????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ????? ?????

Подписано к печати ??.??.201? г. Формат $60 \times 90/16$. Печать офсетная.
Объем ?? печ. л. Тираж ?000 экз. Заказ № .

Отпечатано ?????????????

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга», Москва, Большой Власьевский пер., д. 11. Тел. (495) 745–80–31.
E-mail: biblio@mccme.ru

ISBN 978-5-4439-????-?

Foreword

The present booklet is a compilation of the lecture notes given as handouts to students taking the Topology-I course that I taught in the fall semester of 2015 in the framework of the Math in Moscow program. Actually, it is a rewritten version of the booklet of the same title, jointly authored by Victor Prasolov and myself. The sequence of lectures remains almost the same, the exercises are practically unchanged, but the exposition within each lecture has been severely revised, new figures have been added.

The resulting course is a one-semester introduction to topology, emphasizing the geometric and algebraic aspects. What material will be covered? We shall give a brief answer to that question here, together with a few comments about why we chose that particular material.

In this course, we work mainly with classical subsets of Euclidean spaces (graphs, surfaces, polyhedra, CW-spaces, etc.) rather than with abstract topological spaces. We do not strive for maximal generality, because we regard Topology more as a tool (used in other mathematical disciplines) than an object of study for its own sake. We do introduce the notion of topological space (in Lecture 2) and prove the basic theorems of general topology (after having proved them in the particular case of subsets of Euclidean space in Lecture 1), but we do not go deeply into the theory. We follow that up (Lecture 3) with the main constructions used in topology (Cartesian product, quotient space, wedge, join, cone, suspension, simplicial and CW-spaces, etc.).

The first topological object that we study thoroughly are (triangulable) surfaces (Lectures 4 and 5). We classify them up to homeomorphism, using the above-mentioned constructions and the first serious invariant in the course—the Euler characteristic.

Next (Lecture 6), we introduce the notion of homotopy, which plays a deciding role in the evolution of topology towards algebra, from "point set topology" to "homotopical topology" (aka "algebraic topology"). Here the second important invariant in our course, the degree of circle maps, makes its appearance.

This invariant is immediately applied outside of topology, to the theory of vector fields (which are not topological objects, they live in the theory of differential equations). This is done in Lectures 7 and 8, which culminate in the beautiful Poincaré–Hopf theorem, proved here for surfaces. In the proof, two invariants—the index of a vector field (defined via the degree of circle maps) and the Euler characteristic—unexpectedly come together.

The remaining topological objects studied in the course are plane curves, covering spaces, and knots. In the study of plane curves (Lecture 9), two more invariants appear: the winding number (also defined via the degree of circle maps), which allows to prove the Whitney–Graustein theorem on the classification of regular curves, and the degree of a point with respect to a curve, which is the simplest example of a finite type invariant (in the sense of Vassiliev), and is used here to prove the so-called "fundamental theorem of algebra" (algebraic equations always have roots).

The study of covering spaces is precluded by the introduction (Lecture 10) of the fundamental group $\pi_1(X)$, whose main role in this course is to show how an algebraic object can almost entirely govern a complicated geometric situation, and allows to prove deep facts about covering spaces my means of short and simple algebraic arguments. But before using $\pi_1(X)$ to classify covering spaces (Lecture 11), we use it to prove the Brouwer fixed point theorem in dimension two. For the students, this is a first occasion to come in contact with the category theory language: the functoriality of the fundamental group is precisely what reduces the proof of Brouwer's deep geometric theorem to almost trivial algebra.

The final (12-th) lecture is a brief survey of knot theory, a classical branch of topology which experienced a striking revival at the end of the 20-th century. After introducing the geometry, arithmetic, and combinatorics of knots and links, we show how the Conway axioms can be used to compute the Alexander–Conway polynomial of knots, thus introducing the students "diagrammatic combinatorics", a new type of mathematical argumentation, particularly popular today in mathematical physics.

In accord with our general philosophy—topology as a tool, study of concrete objects—the ability of using topological methods to solve problems is essential. Solving the problems, much more than memorizing the theory, is the way to really master topology.

* * *

This booklet would never have been published without the help of a number of colleagues and friends. I am particularly grateful to Victor Prasolov, who was the leading force in the development of this course, to Mikhail Panov, who produced most of the illustrations, to Serge Lvovsky for careful and precise editing, to Victor Shuvalov for reformatting the original TEX file. I am also grateful to the numerous students who took the course, pointed out errors and showed, by their reactions, whether or not the exposition was adequate.

Lecture 1

The topology of subsets of $\mathbb{R}^n$

The basic material of this lecture should be familiar to you from Advanced Calculus courses, but we shall revise it in detail to ensure that you are comfortable with its main notions (the notions of open set and continuous map) and know how to work with them.

1.1. Continuous maps

"Topology is the mathematics of continuity"

Let $\mathbb{R}$ be the set of real numbers. A function $f\colon \mathbb{R} \to \mathbb{R}$ is called *continuous at the point* $x_0 \in \mathbb{R}$ if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that the inequality

$$|f(x_0) - f(x)| < \varepsilon$$

holds for all $x \in \mathbb{R}$ whenever $|x_0 - x| < \delta$. The function f is called *continuous* if it is continuous at all points $x \in \mathbb{R}$.

This is basic one-variable calculus.

Let $\mathbb{R}^n$ be n-dimensional space. By $O_r(p)$ denote the *open ball* of radius $r > 0$ and center $p \in \mathbb{R}^n$, i.e., the set

$$O_r(p) := \{q \in \mathbb{R}^n : d(p, q) < r\},$$

where d is the distance in $\mathbb{R}^n$. A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called *continuous at the point* $p_0 \in \mathbb{R}^n$ if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $f(p) \in O_\varepsilon(f(p_0))$ for all $p \in O_\delta(p_0)$. The function f is called *continuous* if it is continuous at all points $p \in \mathbb{R}^n$.

This is (more advanced) calculus in several variables.

A set $G \subset \mathbb{R}^n$ is called *open in* $\mathbb{R}^n$ if for any point $g \in G$ there exists a $\delta > 0$ such that $O_\delta(g) \subset G$. Let $X \subset \mathbb{R}^n$. A subset $U \subset X$ is called *open in* X if for any point $u \in U$ there exists a $\delta > 0$ such that $O_\delta(u) \cap X \subset U$.

An equivalent property: $U = V \cap X$, where V is an open set in $\mathbb{R}^n$. Clearly, any union of open sets is open and any finite intersection of open sets is open. Let X and Y be subsets of $\mathbb{R}^n$. A map $f\colon X \to Y$ is called *continuous* if the preimage of any open set is an open set, i.e.,

$$V \text{ is open in } Y \implies f^{-1}(V) \text{ is open in } X.$$

This is basic topology.

Let us compare the three definitions of continuity. Clearly, the topological definition is not only the shortest, but is conceptually the simplest. Also, the topological definition yields the simplest proofs. Here is an example.

Theorem 1.1. *The composition of continuous maps is a continuous map. In more detail, if X, Y, Z are subsets of $\mathbb{R}^n$, $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous maps, then their composition, i.e., the map $h = g \circ f\colon X \to Z$ given by $h(x) := g(f(x))$, is continuous.*

Proof. Let $W \subset Z$ be open. Then the set $V := f^{-1}(W) \subset Y$ is open (because f is continuous). Therefore, the set $U := g^{-1}(V) \subset X$ is open (because g is continuous). But $U = h^{-1}(W)$. □

Compare this proof with the proof of the corresponding theorem in basic calculus. This proof is much simpler.

The notion of open set, used to define continuity, is fundamental in topology. Other basic notions (neighborhood, closed set, closure, interior, boundary, compactness, path connectedness, etc.) are defined by using open sets.

1.2. Closure, boundary, interior

By a *neighborhood* of a point $x \in X \subset \mathbb{R}^n$ we mean any open set (in X) that contains x.

Let $A \subset X$; an *interior point* of A is a point $x \in A$ which has a neighborhood U in X contained in A. The set of all interior points of A is called the *interior* of A in X and is denoted by $\mathrm{Int}(A)$. An *isolated point* of A in X is a point $a \in A$ which has a neighborhood U in X such that $U \cap A = a$.

A *boundary point* of A in X is a point $x \in X$ such that any neighborhood $U \ni x$ in X contains points of A and points not in A, i.e., $U \cap A \neq \varnothing$ and $U \cap (X - A) \neq \varnothing$; the boundary of A is denoted by $\mathrm{Bd}(A)$ or ∂A. The union of A and all the boundary points of A is called the *closure* of A in X

and is denoted by $\mathrm{Clos}(A, X)$ (or $\mathrm{Clos}(A)$, or $\overline{A}$, if X is clear from the context).

Theorem 1.2. *Let $A \subset \mathbb{R}^n$.*

(a) *A is closed if and only if it contains all of its boundary points.*

(b) *The interior of A is the largest (by inclusion) open set contained in A.*

(c) *The closure of A is the smallest (by inclusion) closed set containing A.*

(d) *The boundary of a set A is the difference between the closure of A and the interior of A:* $\mathrm{Bd}(A) = \mathrm{Clos}(A) - \mathrm{Int}(A)$.

The proofs follow directly from the definitions, and you should remember them from the Calculus course. You should be able to write them up without much trouble in the exercise class.

1.3. Topological equivalence

"A topologist is person who can't tell the difference between a coffee cup and a doughnut."

The goal of this section is to teach you to visualize objects (geometric figures) the way topologists see them, i.e., by regarding figures as equivalent if they can be bijectively deformed into each other. This is something you have not been taught to do in calculus courses, and it may take you some time before you will become able to do it.

Let X and Y be "geometric figures," i.e., arbitrary subsets of $\mathbb{R}^n$. Then X and Y are called *topologically equivalent* or *homeomorphic* if there exists a *homeomorphism* of X onto Y, i.e., a continuous bijective map $h\colon X \to Y$ such that the inverse map h^{-1} is continuous.

For the topologist, homeomorphic figures are the same figure: a circle is the same as the boundary of a square, or that of a triangle, of a hexagon, of an ellipse; an arc of a circle is the same as a closed interval, a 2-dimensional disk is the same as the square, or as a triangle together with its inner points; the boundary of a cube is the same as a sphere, or as the boundary of a cylinder, or (the boundary of) a tetrahedron.

If a property does not change under any homeomorphism, then this property is called *topological.* Examples of topological properties are compactness and path connectedness (they will be defined later in this lecture). Examples of properties that are *not* topological are length, area, volume, and boundedness. The fact that boundedness is not a topological property may seem rather surprising; as an illustration, we shall prove that

the open interval $(0, 1)$ *is homeomorphic to the real line* $\mathbb{R}$ (!)

This is proved by constructing an explicit homeomorphism $h\colon (0, 1) \to \mathbb{R}$ as the composition of the two homeomorphisms p and s shown in Figure 1.1.

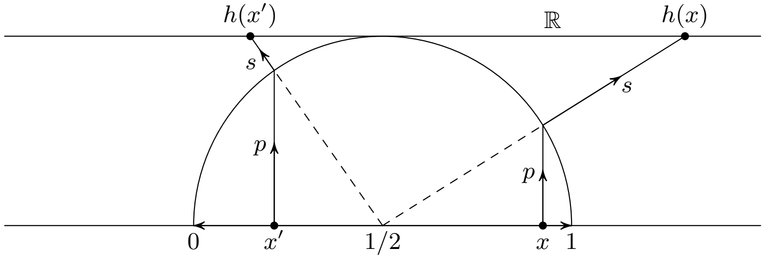


Figure 1.1. The homeomorphism $h\colon (0, 1) \to \mathbb{R}$

For another illustration, look at Figure 1.2; you should intuitively feel that the torus is not homeomorphic to the sphere (although we are at present unable to prove this!). However, the ordinary torus *is* homeomorphic to the knotted torus in the figure, although they look "topologically very different"; they provide examples of figures that are homeomorphic, but are embedded in $\mathbb{R}^3$ in different ways. We shall come back to this distinction later in the course, in particular in the lecture on knot theory.

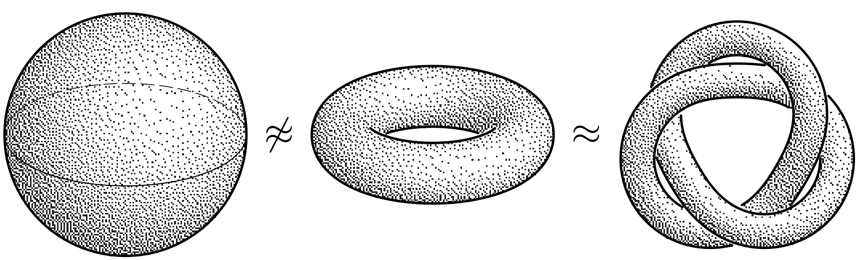


Figure 1.2. The sphere and two tori

We conclude this lecture by studying two basic topological properties of geometric figures that will be constantly used in this course.

1.4. Path connectedness

A set $X \subset \mathbb{R}^n$ is called *path connected* if any two points of X can be joined by a path, i.e., if for any $x, y \in X$ there exists a continuous map $\varphi\colon [0, 1] \to X$ such that $\varphi(0) = x$ and $\varphi(1) = y$.

Theorem 1.3. *The continuous image of a path connected set is path connected. In more detail, if the map $f\colon X \to Y$ is continuous and X is path connected, then $f(X)$ is path connected.*

Proof. Let $y_1, y_2 \in f(X)$. Let $X_1 := f^{-1}(y_1)$ and $X_2 := f^{-1}(y_2)$. Let x_1 and x_2 be arbitrary points of X_1 and X_2, respectively. Then there exists a continuous map $\varphi\colon [0, 1] \to X$ such that $\varphi(0) = x_1$ and $\varphi(1) = x_2$ (because X is path connected). Let $\psi\colon [0, 1] \to f(X)$ be defined by $\psi := f \circ \varphi$. Then ψ is continuous (by Theorem 1.1), $\psi(0) = y_1$ and $\psi(1) = y_2$. $\square$

Thus we have shown that path connectedness is a topological property.

1.5. Compactness

A family $\{U_\alpha\}$ of open sets in $X \subset \mathbb{R}^n$ is called an *open cover* of X if this family covers X, i.e., if $\bigcup_\alpha U_\alpha \supset X$. A *subcover* of $\{U_\alpha\}$ is a subfamily $\{U_{\alpha_\beta}\}$ such that $\bigcup_\beta U_{\alpha_\beta} \supset X$, i.e., the subfamily also covers X. The set X is called *compact* if every open cover of X contains a finite subcover.

Note the importance of the word "every" in the last definition: a set in noncompact if *at least one* of its open covers contains no finite subcover of X. As an illustration, let us show that

the open interval $(0, 1)$ *is not compact.*

Indeed, this follows from the fact that any finite subfamily of the cover $\{U_1, U_2, \dots\}$ shown in Figure 1.3 obviously does not cover $(0, 1)$.

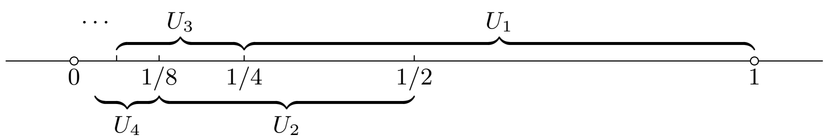


FIGURE 1.3. The open interval $(0, 1)$ is not compact

Theorem 1.4. *The continuous image of a compact set is compact, i.e., if a map $f\colon X \to Y$ is continuous and $X \subset \mathbb{R}^n$ is compact, then $f(X)$ is compact.*

Proof. Let $\{V_\alpha\}$ be an open covering of $f(X)$. Then each $U_\alpha := f^{-1}(V_\alpha)$ is open in X (by the definition of continuity) and so $\{U_\alpha\}$ is an open covering of X. But X is compact, hence $\{U_\alpha\}$ has a finite subcovering, say $\{U_{\alpha_1}, \dots, U_{\alpha_N}\}$. Then $\{f(U_{\alpha_1}), \dots, f(U_{\alpha_N})\}$ is obviously a finite subcover of $\{V_\alpha\}$. $\square$

Thus we have shown that compactness is a topological property.

Fact 1.5. *A set $X \subset \mathbb{R}^n$ is compact if and only if X is closed and bounded.*

We do not give the proof of this fact because it not really topological: the word "bounded" makes no sense to a topologist; the proof is usually given in calculus courses.

1.6. Exercises

1.1. Using the ε-δ definition of continuity, give a detailed proof of the fact that the composition of two continuous functions is continuous.

1.2. Let $F\colon \mathbb{R}^2 \to \mathbb{R}$. Suppose the functions $f_{1,x_0}(y) := F(x_0, y)$ and $f_{2,y_0}(x) := F(x, y_0)$ are continuous for any $x_0, y_0 \in \mathbb{R}$. Is it true that $F(x, y)$ is continuous?

1.3. Prove the four assertions (a)–(d) of Theorem 1.2.

1.4. The towns A and B are connected by two roads. Two travellers can walk along these roads from A to B so that the distance between them at any moment is less than or equal to 1 km. Can one traveller walk from A to B and the other from B to A (using these roads) so that the distance between them at any moment is greater than 1 km?

1.5. Suppose $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$. The *distance* from the point x to the subset A is equal to $d(x, A) = \inf\{\|x - a\| : a \in A\}$.

(i) Prove that the function $f(x) = d(x, A)$ is continuous for any $A \subset \mathbb{R}^n$.

(ii) Prove that if the set A is closed, then the function $f(x) = d(x, A)$ is positive for any $x \notin A$.

1.6. Let X be the subset of $\mathbb{R}^2$ given by the equation $xy = 0$ (X is the union of two lines). Give some examples of neighborhoods: (a) of the point $(0, 0)$; (b) of the point $(0, 1)$.

1.7. Describe the set of points x in $\mathbb{R}^2$ such that $d(x, A) = 1; 2; 3$, where the set A is given by the formula:

(a) $x^2 + y^2 = 0$; (b) $x^2 + y^2 = 2$;
(c)* $x^2 + 2y^2 = 2$; (d) the square of area two.

1.8. Let A and B be two subsets of the set X that was defined in Exercise 1.6. Suppose that A and B are homeomorphic and A is open in X. Is it true that B is also open in X?

1.9. Construct a homeomorphism between the boundary of the cube $\mathbb{I}^3$ and the sphere $\mathbb{S}^2$.

1.10. Construct a homeomorphism between the plane $\mathbb{R}^2$ and the open disk $\mathbb{B}^2 := \{v \in \mathbb{R}^2 : |v| < 1\}$.

1.11. Construct a homeomorphism between the plane $\mathbb{R}^2$ and the sphere $\mathbb{S}^2$ with one point removed.

Lecture 2

Abstract topological spaces

In this lecture, we move from the topological study of concrete geometrical figures (subsets of $\mathbb{R}^n$) to the axiomatic study of abstract topological spaces. What is remarkable about this approach is the simplicity of the underlying axioms (based on the notion of open set, now an undefined concept in the axiomatics), which nevertheless allow to generalize the deep theorems about subsets of $\mathbb{R}^n$ (proved in the previous lecture) to subsets of any abstract topological space, by reproducing the proofs practically word for word.

2.1. Topological spaces

By definition, an (*abstract*) *topological space* $(X, \mathcal{T} = \{U_\alpha\})$ is a set X of arbitrary elements $x \in X$ (called *points*) and a family $\mathcal{T} = \{U_\alpha\}$ (called the *topology of the space* X) of subsets of X (called *open sets*) such that

(1) X and $\varnothing$ are open;

(2) if U and V are open, then $U \cap V$ is open;

(3) if $\{V_\beta\}$ is any collection of open sets, then the set $\bigcup_\beta V_\beta$ is open.

Any set $X \subset \mathbb{R}^n$ is a topological space if the family of open sets is defined as in Section 1.1. (The proof is a straightforward exercise.) All the definitions from Sections 1.2–1.4 are valid for any topological space (and not only for subsets of $\mathbb{R}^n$), because they only use the notion of open set. All the theorems (and their proofs) from the previous lecture are also valid. At this point the reader should read through these proofs again and check that, indeed, only the properties of open sets appearing in the axioms are used.

In order to define a topological space, we don't have to specify *all* the open sets: there is a more "economical" way of defining the topology. For a topological space $(X, \mathcal{T})$, we say that a subset $\mathcal{T}_0 \subset \mathcal{T} = \{U_\alpha\}$) is a *base* *of the topology* of $(X, \mathcal{T})$ if for any open set $U \in \mathcal{T}$ there exists a collection $\{V_\beta\}$ of open sets in $\mathcal{T}_0$ such that $U = \bigcup_\beta V_\beta$.

Clearly, any base of the topology uniquely determines the whole topology (how?). For example, the set of all open balls in $\mathbb{R}^n$ is a base of the standard topology of Euclidean space.

Examples. (1) Any set D becomes a topological space if it is supplied with the *discrete topology*, i.e., if any set is declared open. Obviously, a topology is discrete if and only if any point is an open set.

(2) Any set X supplied with only two open sets (the empty set and X itself) is a topological space with the *trivial topology*.

(3) Any metric space M (see the definition in the next section) is a topological space in the *metric topology*, which is given by the base of all open balls $O_r(m) := \{m' \colon d(m', m) < r\}$ in M, where d is the distance function in M.

(4) The space $\mathcal{C}[0, 1]$ of continuous real-valued functions on the closed interval $[0, 1] \subset \mathbb{R}$ has a standard topology given by the base of open balls $O_r(f) := \{g \colon \sup_x(|g(x) - f(x)|) < r\}$.

Many more nontrivial examples will be given at the end of this lecture, in the exercise class and in subsequent lectures.

2.2. Metric spaces

A *metric space* is a set M supplied with a *metric* (or *distance function*), i.e., a function $d \colon M \times M \to \mathbb{R}$ such that

(1) for all $x, y \in M$, $d(x, y) \geqslant 0$ (*nonnegativity*);
(2) for all $x, y \in M$, $d(x, y) = 0$ iff $x = y$; (*identity*);
(3) for all $x, y \in M$, $d(x, y) = d(y, x)$ (*symmetry*);
(4) for all $x, y, z \in M$, $d(x, z) \leqslant d(x, y) + d(y, z)$ (*triangle inequality*).

The most popular example of a metric space is Euclidean space $\mathbb{R}^n$ (and its subsets) with the standard metric:

$$d(p, q) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \quad \text{where } p = (x_1, \ldots, x_n),\ q = (y_1, \ldots, y_n).$$

Other less familiar examples will appear in the exercise classes.

As we mentioned above, any metric space (M, d) becomes a topological space in the metric topology. Conversely, it is *not* true that any topological space $(X, \mathcal{T})$ has a metric (i.e., possesses a distance function for which the metric topology coincides with $\mathcal{T}$). Until the middle of the 20th century one of the main problems of topology was to find necessary and sufficient conditions for a topological space $(X, \mathcal{T})$ to be *metrizable*, i.e., for X to have a metric such that the corresponding metric topology coincides with $\mathcal{T}$.

2.3. Induced topology

If A is a subset of a topological space X, then A acquires a topological structure in a natural way: the topology on A is *induced* from X if we declare all the intersections of open sets of X with A to be the open sets of A. It is easy to check that A with the induced topology is indeed a topological space (i.e., satisfies axioms (1)–(3) from Section 2.1).

It is important to note that open sets in the induced topology of A are not necessarily open in X (in fact, in most cases they are not).

Whenever we consider a subset of a topological space, we will always regard it as a topological space in the induced topology without explicit mention. Speaking of open sets, however, one should always make clear with respect to what set or subset openness is understood. Thus the open interval $(0, 1)$ is open on the real line, but not in the plane.

2.4. Connectedness

In the previous lecture, we defined path connectedness of subsets of $\mathbb{R}^n$; that definition remains valid, word for word, for topological spaces. Intuitively, pathconnectedness of a topological space means that you can move continuously within the space from any point to any other point. But there is another definition of connectedness based on the idea that a connected set is "a set that consists of one piece". The rigorous formalization of the idea of "consisting of one piece" is as follows.

A topological space X is called *connected* if it is not the union of two open, closed, nonempty, and nonintersecting sets, i.e., $X = A \cup B$, where A and B are both open, closed, and nonempty, implies $A \cap B \neq \varnothing$.

What is the relationship between the notions of connectedness and path connectedness?

Theorem 2.1. *Any path connected topological space is connected, but there exist connected topological spaces that are not path connected.*

Proof. Suppose that the space X is path connected. Arguing by contradiction, let us assume that it is the disjoint union of two open and closed nonempty sets A and B. Let $a \in A$, $b \in B$. Then there exists a continuous map $f\colon [0,1] \to X$ such that $f(0)=a$ and $f(1)=b$. Denote $A_0 := f^{-1}(A)$ and $B_0 := f^{-1}(B)$. These two sets are disjoint, open (as inverse images of open sets) and cover the closed interval $[0,1]$ (because $f([0,1]) \subset X = A \cup B$). We know that $1 \in B_0$. Let ξ be the least upper bound of A_0. If $\xi \in A_0$, then A_0 cannot be open, so ξ belongs to B_0; but then B_0 cannot be open. A contradiction.

Concerning the converse statement, see Exercise 2.12. □

Connectedness, like path connectedness, is not only a topological property—it is preserved by *any* continuous maps (not only by homeomorphisms).

Theorem 2.2. *The continuous image of a connected set is connected, i.e., if a map $f\colon X \to Y$ is continuous and X is connected, then $f(X)$ is connected.*

Proof. We argue by contradiction: suppose that X is connected, but $f(X)$ is not. Then $f(X)=A\cup B$, where both A and B are both closed and open, and don't intersect. Denote $A'=f^{-1}(A)$ and $B'=f^{-1}(B)$. Then $X=A\cup B$, $A\cap B=\varnothing$, both A and B are open (as preimages of open sets) and closed (as complements to open sets). But this means that X is not connected—a contradiction. □

Roughly speaking, a connected component of a nonconnected set is just one of its many "pieces". The formal definition is this: a *connected component* of a not necessarily connected space X is any connected subset of X not contained in a larger connected subset of X. It is easy to prove that any connected component of a space X is both open and closed in X.

2.5. Separability

An important type of property for topological spaces comes from various separability axioms, which specify how well it is possible to "separate" points and/or sets (i.e., put them into nonintersecting neighborhoods). We only define one such property, the most natural and classical one: a topological space is said to be a *Hausdorff space* if any two distinct points possess nonintersecting neighborhoods. Obviously, Euclidean space and any of its subsets are Hausdorff, as are indeed any metric spaces (why?). The sad fact that there exist non-Hausdorff spaces will be considered in the exercise class.

2.6. More examples of topological spaces

In this section, we list twelve classical mathematical objects (not necessarily familiar to you) coming from completely different areas of mathematics. All of them are topological spaces. In the exercise class (and in doing the homework assignments), you will learn how to define their topology (by introducing an appropriate base). You will perhaps be surprised to learn that certain objects from different parts of mathematics and physics, which at first glance have nothing in common, turn out to be topologically equivalent (homeomorphic).

We begin with examples coming from algebra.

(1) The group $\mathrm{Mat}(n, n)$ of all nondegenerate $n \times n$ matrices.

(2) The group $\mathrm{O}(n)$ of all orthogonal transformations of $\mathbb{R}^n$.

(3) The set of all polynomials of degree n with leading coefficient 1.

The next examples come from geometry.

(4) The real projective space $\mathbb{R}P^n$ of dimension n.

(5) The Grassmanian $G(k, n)$, i.e., the set of k-dimensional planes containing the origin in n-dimensional affine space.

(6) The hyperbolic plane.

The next example comes from complex analysis.

(7) The Riemann sphere $\overline{\mathbb{C}}$ and, more generally, Riemann surfaces.

Here are some examples from classical mechanics.

(8) The configuration space of a solid rotating about a fixed point in 3-space.

(9) The configuration space of a rectilinear rod rotating in 3-space about (a) one of its extremities, (b) its midpoint.

Here are two from algebraic geometry.

(10) The set of solutions $p = (x_1, \dots, x_9) \in \mathbb{R}^9$ of the following system of 6 equations:

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= 1, & x_1x_4 + x_2x_5 + x_3x_6 &= 0, \\ x_4^2 + x_5^2 + x_6^2 &= 1, & x_1x_7 + x_2x_8 + x_3x_9 &= 0, \\ x_7^2 + x_8^2 + x_9^2 &= 1, & x_4x_7 + x_5x_8 + x_6x_9 &= 0. \end{aligned}$$

(11) Any affine variety in the *Zariski topology* is a topological space.

In conclusion, an example from dynamical systems (differential equations).

(12) The phase space of billiards on the disk.

2.7. Exercises

2.1. Prove that any constant map is continuous.

2.2. For any subsets $A, B \subset \mathbb{R}^n$, define the *distance* between A and B by putting $d(A, B) := \inf\{\|a - b\| : a \in A, b \in B\}$.

(a) Is it true that $d(A, C) \leqslant d(A, B) + d(B, C)$?

(b) Let $A \subset \mathbb{R}^n$ be a closed subset, let $C \subset \mathbb{R}^n$ be a compact subset. Prove that there exists a point $c_0 \in C$ such that $d(A, C) = d(A, c_0)$. Further, prove that if the set A is also compact, then there exists a point $a_0 \in A$ such that $d(A, C) = d(a_0, c_0)$.

2.3. Prove that any closed subspace of a compact space is compact.

2.4. Prove that the topology of $\mathbb{R}^n$ has a countable base (i.e., a base consisting of a countable family of open sets).

2.5. Introduce a "natural" topology on

(a) the set $\mathrm{Mat}(m, n)$ of matrices of size $n \times m$;

(b) the real projective space $\mathbb{R}P(n)$ of dimension n;

(c) the Grassmannian $G(k, n)$, i.e., the set of k-dimensional planes containing the origin of n-dimensional affine space;

(d) the set of solutions $p = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ of the following system of two equations: $x_1^2 + x_2^3 + x_3^4 + x_4^5 = 1$ and $x_1 x_2 x_3 x_4 = -1$;

(e) the set of all polynomials of degree n with leading coefficient 1.

2.6. (a) Is the topological space $\mathrm{GL}(n)$ connected?

(b) Prove that the topological space $\mathrm{SO}(3)$ is connected.

(c) Prove that the topological space $\mathrm{GL}(3)$ consists of two connected components.

2.7. (a) Prove that the function $d(x, y) = \max\{|x_i - y_i|, i = 1, \dots, n\}$ where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ defines a metric in $\mathbb{R}^n$.

(b) Prove that $d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ is a metric in $\mathbb{R}^n$.

(c) Draw some ε-neighborhood of the point $(0, 0, \dots, 0)$ in the metrics defined in (a) and (b).

2.8. Prove that any metric space is Hausdorf and construct an example of a non-Hausdorff space.

2.9. Let X be a Hausdorff space. Prove that for any two distinct points $x, y \in X$ there exists a neighborhood $U \ni x$ such that its closure does not contain the point y.

2.10. Let C be a compact subspace of a Hausdorff space X. Let $x \in X \setminus C$. Prove that the point x and the set C have disjoint neighborhoods.

2.11. Prove that any two disjoint compact subsets of a Hausdorff space have disjoint (open) neighborhoods.

2.12. Give an example of a connected topological space which is not path connected.

Lecture 3

Topological constructions

In this lecture, we study the basic constructions used in topology. These constructions transform one or several given topological spaces into a new topological space. Starting with the simplest topological spaces and using these constructions, we can create more and more complicated spaces, including those which are the main objects of study in topology.

3.1. Disjoint union

The *disjoint union* of two topological spaces X and Y, in the case when the two sets X and Y do not intersect, is the union of the sets X and Y with the following topology: a set W in $X \cup Y$ is open if the sets $W \cap X$ and $W \cap Y$ are open in X and Y, respectively; if the two sets X and Y intersect, the definition is a little trickier: first we artificially make them nonintersecting by considering, instead of the set Y, the same set of elements but marked, say, with a star, i.e., $Y^* := \{(y, *) \colon y \in Y\}$, and then proceed as before, declaring that a set W in $X \cup Y^*$ is open if the sets $W \cap X$ and $W \cap Y^*$ are open in X and Y^*, respectively. In both cases, we obtain a topological space denoted by $X \sqcup Y$.

This choice of topology ensures that both natural inclusions $X \hookrightarrow X \cup Y$ $(x \mapsto x)$ and $Y \hookrightarrow X \cup Y$ $(y \mapsto y)$ are continuous maps.

It is easy to see that the subsets X and Y (we do not explicitly write the stars (if any) in Y^*, but consider them implicitly present) are both open and closed in $X \sqcup Y$, so that the set $X \sqcup Y$ is not connected (provided both X and Y are nonempty).

3.2. Cartesian product

Roughly speaking, the Cartesian product of two spaces is obtained by putting a copy of one of the spaces at each point of the other space.

More precisely, let X and Y be topological spaces; consider the set of pairs $X \times Y = \{(x, y) : x \in X, y \in Y\}$ and make $X \times Y$ into a topological space by defining its base: a set $W \subset X \times Y$ belongs to the base if it has the form $W = U \times V$, where U is an open set in X and V is open in Y. It is easy to check that in this way we obtain a topological space, which is called the *Cartesian product* of the spaces X and Y.

This choice of topology ensures that both natural projections $X \times Y \to \to X$ $((x, y) \mapsto x)$ and $X \times Y \to Y$ $((x, y) \mapsto y)$ are continuous maps.

Classical examples: (i) the Cartesian product of two closed intervals is the square; (ii) the Cartesian product of two circles is the torus; (iii) the Cartesian product of two real lines $\mathbb{R}$ is the plane $\mathbb{R}^2$.

Theorem 3.1. *The Cartesian product of the n-disk and the m-disk is the $(n+m)$-disk. The Cartesian product of $\mathbb{R}^n$ and $\mathbb{R}^m$ is $\mathbb{R}^{n+m}$.*

The proof is absolutely straightforward.

3.3. Quotient spaces

Roughly speaking, a quotient space is obtained from a given space by identifying the points of certain subsets of the given space ("dividing" our space by these subsets).

More precisely let X be a topological space and let $\sim$ be an equivalence relation on the set X; we then consider the equivalence classes with respect to this relation as points of the quotient set $X/_\sim$ and introduce a topology in this set by declaring open any subset $U \subset X/_\sim$ such that $U^* := \{x \in \xi : \xi \in U\}$ is open in X. The topological space thus obtained is denoted by $X/_\sim$.

This choice of topology ensures that the natural projection $X \to X/_\sim$ $(x \mapsto \xi_\beta$, where $\xi_\beta \ni x)$ is a continuous map.

Suppose X and Y are topological spaces, A and B are closed subspaces of X and Y, respectively, and $f \colon A \to B$ is a continuous map. (The particular case in which f is a homeomorphism is often considered.) In the disjoint union of X and Y, we identify all points of each set in the family

$$\mathcal{F}_b := \{b \sqcup f^{-1}(b) : b \in B\}.$$

Then we denote the quotient space $(X \cup Y)/_\sim$, where $\sim$ is the equivalence relation identifying points in each of the sets $\mathcal{F}_b$, $b \in B$, $X \cup_f Y$ and say that this space is obtained by *attaching* (or *gluing*) Y to X along f.

If A is a subset of a topological space X, we denote by X/A the quotient space w. r. t. the equivalence relation $x \sim y$ iff $x, y \in A$. For example, we have $\mathbb{D}^n/\partial\mathbb{D}^n \approx \mathbb{S}^n$.

3.4. Cone, suspension, and join

(i) Roughly speaking, the cone over a space is obtained by joining a fixed point by line segments with all the points of the space. More precisely, let X be a topological space; consider the Cartesian product $X \times [0, 1]$ (called the *cylinder* over X) and on it, the equivalence relation $(x, 1) \sim (y, 1)$ for any $x, y \in X$; we define the *cone over* X as the quotient space of the cylinder by the equivalence relation $\sim$:

$$C(X) := (X \times [0, 1])/_{\sim}.$$

Note that all the points (x, t) with $t = 1$ are identified into one point, called the *vertex* of the cone. By definition, the cone over the empty set is one point. The cone over a point is a line segment, the cone over the circle is homeomorphic to the disk (although it is more natural to think of it as the lateral surface of the ordinary circular cone).

(ii) Roughly speaking, the suspension over a topological space is obtained by joining two fixed points by segments with all the points of the given space. Another heuristic way of saying this is that the suspension is a double cone (on "different sides") over that space.

More precisely, let X be a topological space; consider the Cartesian product $X \times [-1, 1]$ and on it, the equivalence relation

$$(x, 1) \approx (y, 1) \quad \text{and} \quad (x, -1) \approx (y, -1)$$

for any $x, y \in X$; now define the *suspension over* X as the quotient space of the cylinder $X \times [-1, 1]$ by the equivalence relation $\approx$:

$$\Sigma(X) := (X \times [-1, 1])/_{\approx}.$$

By definition, the suspension over the empty set is the two point set $\mathbb{S}^0$. The suspension over the two point set is homeomorphic to the circle, that over the circle is homeomorphic to the 2-sphere.

The notion of suspension is extremely important in topology, particularly in algebraic topology (surprisingly, it is much more important than that of the cone).

(iii) Roughly speaking, the join of two spaces is obtained by joining each pair of points from the two spaces by a segment.

More precisely, suppose that X and Y are topological spaces; consider the Cartesianproduct $X \times [-1, 1] \times Y$ and identify (via an equivalence relation that will be denoted by $\equiv$) all pairs of points of the form $(x_1, 1, y) \equiv (x_2, 1, y)$ as well as all pairs of the form $(x, -1, y_1) \equiv (x, -1, y_2)$. The topological space $X * Y$ thus obtained,

$$X * Y := (X \times [-1, 1] \times Y)/_{\equiv},$$

is called the *join* of the spaces X and Y.

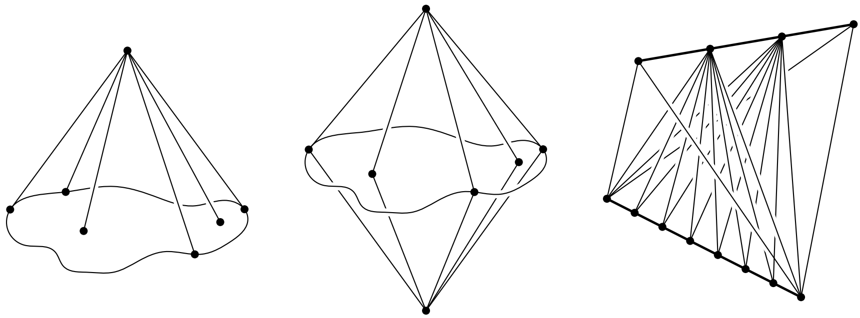


FIGURE 3.1. Cone and suspension. Join of two closed intervals

Theorem 3.2. *The cone over the n-sphere is the $(n + 1)$-disk and the cone over the n-disk is the $(n + 1)$-disk. The suspension over the n-sphere is the $(n + 1)$-sphere and the suspension over the n-disk is the $(n+1)$-disk. The join of the n-disk and the m-disk is the $(n+m+1)$-disk. The join of the n-sphere and the m-sphere is the $(n + m + 1)$-sphere.*

The proof is not difficult: one performs the construction in a Euclidean space of the appropriate dimension; in each case the corresponding homeomorphism is not hard to construct, although for large values of n and m it is difficult to visualize. The simplest (and only really "visual") nontrivial example is the join of two segments (which is the tetrahedron, otherwise known as the 3-simplex); it is shown in Figure $\overline{3}$.1.

3.5. Simplicial spaces

A 0-simplex is a point, a 1-simplex is a closed interval, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on. More generally and precisely, we define an *n-dimensional simplex* σ_n (*n-simplex* for short) as a topological space supplied with a homeomorphism

$$h\colon \sigma_n \to \Delta^n = [e_0, e_1, \dots, e_n],$$

where Δ^n is the convex hull of the set of $n+1$ points consisting of the origin $0=e_0$ and the endpoints $e_1, \ldots, e_n$ of the basis unit vectors of Euclidean space $\mathbb{R}^n$. The n-simplex is of course homeomorphic to the n-disk $\mathbb{D}^n$, but it has a richer structure coming from the homeomorphism h. Namely, for any i, $0 \leqslant i \leqslant n$, it has a set of *$i$-faces*, each i-face is the preimage under h of the convex hull in $\mathbb{R}^n$ of i points from the set $\{e_0, e_1, \ldots, e_n\}$. The 0-faces of an n-simplex are called *vertices*, and we often write

$$\sigma_n = [0, 1, \ldots, n],$$

where by abuse of notation i, $i=0, 1, \ldots, n$, denotes the vertex $h^{-1}(e_i)$.

Thus the 3-simplex possesses four 2-faces (triangles), six 1-faces (edges) and four 0-faces (vertices). By convention, we agree that the empty set is regarded as the (-1)-dimensional simplex. Note that the 3-simplex (as well as its faces), inherits a linear structure from $\mathbb{R}^3$ by the homeomorphism $h\colon \sigma_3 \to \Delta^3 \subset \mathbb{R}^3$.

We now define a *finite simplicial space* X (also called *finite simplicial complex*) as the space obtained from the disjoint union of a finite set of simplices by gluing some of their faces together by homeomorphisms; it is assumed that the attaching homeomorphisms respect the linear structure of the faces (so that after the gluing is performed, all the simplices have a coherent linear structure). In this course, we will not consider the more general notion of simplicial space with a possibly infinite number of simplices, and so will often drop the adjective finite when speaking of finite simplicial spaces. By the *dimension* of a simplicial space X we mean the dimension of the simplices of the highest dimension in X and we often write it in the form of a superscript, writing X^n for an n-dimensional simplicial space.

A more geometric way of defining a simplicial space is to represent it as a subset of some Euclidean space, with the simplices being rectilinear geometric subsets of the space. Figure 3.2 shows two such examples of simplicial spaces, represented as lying in $\mathbb{R}^3$: a 2-sphere and a funny 2-dimensional simplicial space.

As the following theorem claims, any finite simplicial space X can be represented as a subset of some Euclidean space $\mathbb{R}^N$ in the sense specified above—one then says that X is *piecewise-linearly embedded* (PL-*embedded* for short) in $\mathbb{R}^N$.

Theorem 3.3. *Any finite n-dimensional simplicial space X^n can be PL-embedded in $\mathbb{R}^{2n+1}$.*

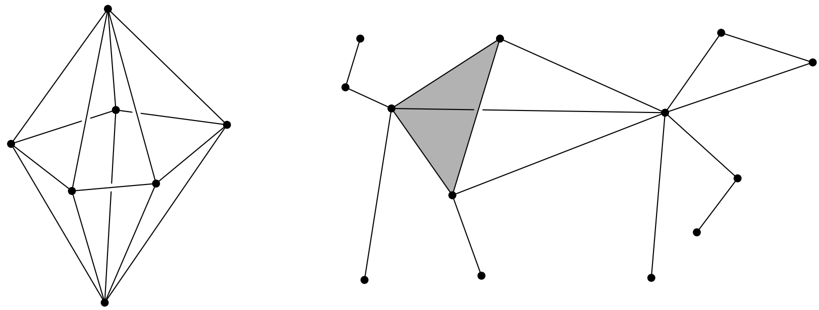


FIGURE 3.2. Two simplicial spaces as subsets of $\mathbb{R}^3$

We shall not use this theorem and therefore omit its proof. The reader may wonder where the exponent $2n+1$ comes from; there are examples of 1-dimensional simplicial spaces (e. g. the so-called $K_{3,3}$ space) that cannot be embedded in $\mathbb{R}^2$.

3.6. CW-spaces

Roughly speaking, a CW-space is a space obtained by inductively attaching k-disks ($k=0, 1, 2, \ldots$) along their boundaries to the $(k-1)$-dimensional part of the previously constructed space via continuous maps of their boundaries (these maps, as well as their images, are called k-cells).

The formal definition of CW-space (also called CW-complex) is the following. Let X be a Hausdorff topological space such that

$$X = \bigcup_{i=0}^{\infty} X^i,$$

where X^0 is a discrete space and the space X^{i+1} is obtained by attaching the disjoint union of $(i+1)$-dimensional closed discs $\bigsqcup_{\alpha\in A} D_\alpha^{i+1}$ to X^i along a continuous map $\bigsqcup_{\alpha\in A} S_\alpha^i \to X^i$, where $S_\alpha^i = \partial D_\alpha^{i+1}$. Let us call the image of D_α^{i+1} and the image of the interior of D_α^{i+1} under the natural map to $X^{i+1} \hookrightarrow X$ *closed cell* and *open cell*, respectively. The space X is called a *CW-space* (or *CW-complex*) if the two following conditions hold:

(C) any closed cell intersects a finite number of open cells;

(W) a set $C \subset X$ is closed iff any intersection of C with a closed cell is closed.

"C" is the abbreviation for "Closure Finite", "W" is the abbreviation for "Weak Topology". If the number of cells is finite, then conditions (C) and (W) hold automatically. Since we will only be considering finite cell spaces in this course, you can forget about conditions (C) and (W).

Note that any simplicial space can be considered as a CW-space (how?). Simplicial spaces are easier to visualize than CW-spaces, because simplices are simpler than cells, but CW-spaces are more economical. For example, the 77-dimensional sphere has a CW-space structure with only two cells, whereas the simplest simplicial structure of that sphere has hundreds of simplices of dimensions $0, 1, 2, \dots, 77$.

3.7. Exercises

3.1. Prove that $\mathbb{D}^n/\partial\mathbb{D}^n \approx \mathbb{S}^n$.

3.2. Prove that the space $\mathbb{S}^1 \times \mathbb{S}^1$ is homeomorphic to the space obtained by the following identification of points of the square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ belonging to its sides: $(x, 0) \sim (x, 1)$ and $(0, y) \sim (1, y)$. (This space is called the torus.)

3.3. Let $I = [0, 1]$. Prove that the space $\mathbb{S}^1 \times I$ is not homeomorphic to the Möbius band.

3.4. Prove that the following spaces (supplied with the natural topology) are homeomorphic:

(a) the set of lines in $\mathbb{R}^{n+1}$ passing through the origin;

(b) the set of hyperplanes in $\mathbb{R}^{n+1}$ passing through the origin;

(c) the sphere $\mathbb{S}^n$ with identified diametrically opposite points (every pair of diametrically opposite points is identified);

(d) the disc $\mathbb{D}^n$ with identified diametrically opposite points of the boundary sphere $\mathbb{S}^{n-1} = \partial\mathbb{D}^n$.

3.5. Prove that the following spaces are homeomorphic:

(a) the set of complex lines in $\mathbb{C}^{n+1}$ passing through the origin;

(b) the sphere $\mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1}$ with identified points of the form λx for every $\lambda \in \mathbb{C}$, $|\lambda| = 1$ (for any fixed point $x \in \mathbb{S}^{2n+1}$);

(c) the disc $\mathbb{D}^{2n} \subset \mathbb{C}^n$ with points of the boundary sphere $\mathbb{S}^{2n-1} = \partial\mathbb{D}^{2n}$ of the form λx for every $\lambda \in \mathbb{C}$, $|\lambda| = 1$ identified for any fixed point $x \in \mathbb{S}^{2n-1}$.

3.6. Prove that $C(\mathbb{D}^n) \approx \mathbb{D}^{n+1}$ and $\Sigma(\mathbb{D}^n) \approx \mathbb{D}^{n+1}$. (Here and below $\approx$ denotes homeomorphisms.)

3.7. Prove that $\mathbb{R}P^1 \approx \mathbb{S}^1$ and $\mathbb{C}P^1 \approx \mathbb{S}^2$.

3.8. Prove that $C(\mathbb{S}^n) \approx \mathbb{D}^{n+1}$ and $\Sigma(\mathbb{S}^n) \approx \mathbb{S}^{n+1}$.

3.9. Is it true (for arbitrary CW-spaces) that (a) $X * Y \approx Y * X$; (b) $(X * Y) * Z \approx X * (Y * Z)$; (c) $C(X * Y) \approx C(X) * Y$; (d) $\Sigma(X * Y) \approx$ $\approx \Sigma(X) * Y$?

3.10. Prove that $\mathbb{S}^n * \mathbb{S}^m \approx \mathbb{S}^{n+m+1}$.

3.11. Prove that $\mathbb{S}^{n+m-1} \setminus \mathbb{S}^{n-1} \approx \mathbb{R}^n \times \mathbb{S}^{m-1}$. (We suppose that the position of $\mathbb{S}^{n-1}$ in $\mathbb{S}^{n+m-1}$ is standard.)

3.12. Prove that (a) the sphere $\mathbb{S}^2$; (b) the torus $\mathbb{T}^2$; (c) the real projective space $\mathbb{R}P^n$; (d) the complex projective space $\mathbb{C}P^n$ are CW-spaces.

3.13. Find an example of a space consisting of cells that satisfies the W-axiom, and does not satisfy the C-axiom and vice versa.

Lecture $\overline{3}$
Graphs

The number of this lecture is overlined, which indicates that the lecture is optional, it should be regarded as additional reading material and a source of problems for the exercise class. In the lecture, we study a very simple class of topological spaces, called graphs. Roughly speaking, a graph G is a set of points, called vertices, some pairs of which are joined by arcs, called edges. Graphs can be defined as purely combinatorial objects, or as topological spaces. Their simplicity is due to the fact that, as combinatorial objects, they are finite and, as topological spaces, they have the smallest nontrivial dimension (one). Nevertheless, they have many surprising, beautiful, and rather intricate properties. We should also note that at the present time graph theory plays a remarkably important role in front-line research in many areas of mathematics.

$\overline{3}$.1. Main definitions

The combinatorial definition of a graph is this: a (*combinatorial*) *graph* G is pair $G=(\mathcal{V},\mathcal{E})$ consisting of a finite set $\mathcal{V}$ of undefined objects, called *vertices*, and a finite collection $\mathcal{E}$ of pairs of vertices, called *edges*; if $e=\{v,v'\}$ is an edge, we say that e *joins* v and v', or that v and v' are the *endpoints* of e; an edge $e=\{v,v'\}$ is said to be a *loop* if $v=v'$; if there are repetitions in the collection of edges $\mathcal{E}$ (i.e., there is more than one edge joining two vertices v and v'), we say that the graph G has *multiple edges.* We shall mostly be studying graphs without loops or multiple edges, and use the term "graph" in that sense; whenever a graph will be allowed to have loops or multiple edges, this will be explicitly mentioned.

Two combinatorial graphs are called *isomorphic* if there exists a bijection between the set of vertices and a bijection between the collection of edges that preserve incidence (i.e., the endpoints of any edge correspond to the endpoints of the corresponding edge).

The topological definition of a graph is this: a (*topological*) *graph* G is topological space G supplied with a finite set $\mathcal{V}$ of distinguished points, called *vertices*, and consisting of the union of a finite number of arcs, called *edges*, each arc being either a broken line joining two vertices or a closed broken line (called a *loop*) containing exactly one vertex; the arcs (including the loops) are assumed pairwise nonintersecting[1]. The graphs considered in this lecture will be subsets of $\mathbb{R}^2$ or $\mathbb{R}^3$ supplied with the topology induced from $\mathbb{R}^2$ or $\mathbb{R}^3$. Unless stated otherwise, we will assume that they contain no loops or multiple edges.

A topological graph is said to be a *realization* of a combinatorial graph if there is a bijection between vertices and a bijection between edges preserving incidence (i.e., endpoints correspond to endpoints); in that situation, we also say that the combinatorial graph is *associated* to the topological one. It is obvious that *two graphs with isomorphic associated combinatorial graphs are homeomorphic*. The converse statement is not true (why?).

The *valency* of a vertex v of a graph G is the number of edges with endpoint v (if there are loops joining v to itself, then each loop contributes 2 to the valency). A *path* joining two vertices v and v' is a sequence of edges of the form $\{v, v_1\}, \{v_1, v_2\}, \ldots, \{v_k, v'\}$; if $v = v'$ and $k > 2$, then the path is called a *cycle*. A graph G is *connected* if any two of its vertices can be joined by a path. A graph is called a *tree* if it is connected and has no cycles; the vertices of valency 1 of a tree are called *leaves*.

Figure $\overline{3}$.1 shows examples of (a) a graph with loops and multiple edges; (b) a tree; (c) a graph without loops or multiple edges, but containing cycles.

The three graphs appearing in the figure are subsets of the plane $\mathbb{R}^2$, but there exist graphs which cannot be placed in the plane. We shall consider them in the next section.

$\overline{3}$.2. Planar and nonplanar graphs

A topological graph is called *planar* if it lies in the plane $\mathbb{R}^2$ (and so its edges have no common internal points). A combinatorial graph G is called *planar* if it can be realized by a planar topological graph.

[1] The assumption that the arcs are polygonal (i.e., are broken lines) is purely technical, it does not restrict (up to topological equivalence) the class of graphs considered, but allows to prove certain statements about embedded graphs which are very difficult to prove without this assumption.

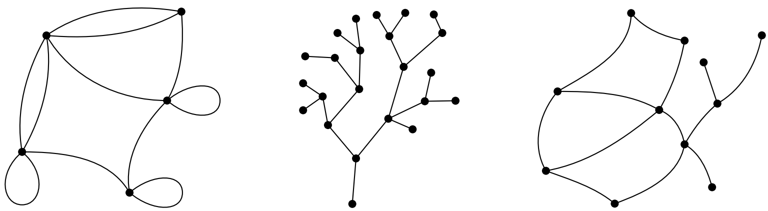


Figure $\overline{3}$.1. Examples of graphs

Denote by K_n the *complete graph on n vertices*, i.e., the graph consisting of n vertices every pair of which is joined by an edge. Denote by $K_{n,m}$ the graph consisting of $n+m$ vertices divided into two parts (n vertices in one part and m vertices in the other), the edges of $K_{n,m}$ joining each pair of vertices from different parts. The figure below represents three examples of the graphs defined above.

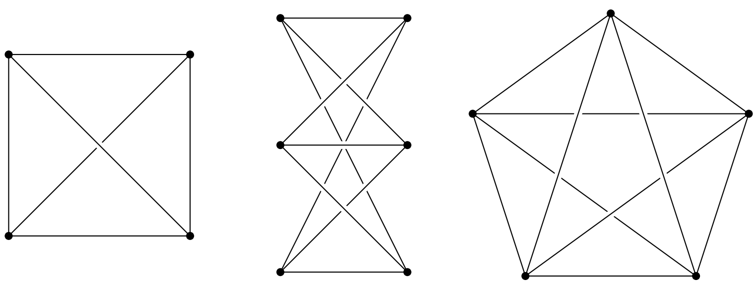


Figure $\overline{3}$.2. The graphs K_4, $K_{3,3}$, and K_5

The three graphs in the figure are pictured as lying in 3-space $\mathbb{R}^3$. Are they planar? The reader will easily draw a graph isomorphic to K_4 embedded in the plane—so K_4 is planar. Attempts to embed the graphs K_5 and $K_{3,3}$ will necessarily fail (the best one can do is to draw a picture of, say, $K_{3,3}$ on the plane with only one pair of edges intersecting).

Theorem $\overline{3}$.1. *The graphs K_5 and $K_{3,3}$ are not planar.*

No simple proof of this beautiful fact is known. In the sections that follow, we shall obtain two different proofs of the theorem. As usual in mathematics, in order to prove that something is impossible (in this case, it is impossible to embed K_5 or $K_{3,3}$ in $\mathbb{R}^2$), we need an invariant. We shall see in subsequent lectures that the invariant that we will use (the Euler characteristic) has many other important applications.

$\overline{3}$.3. Euler characteristic of graphs and planar graphs

If G is a graph (topological or combinatorial), we denote by V_G and E_G the number of vertices and edges of G, respectively; we omit the subscript G if the graph under consideration is clear from the context.

We define the *Euler characteristic of a graph* G by setting

$$\boxed{\chi(G) := V_G - E_G}.$$

Theorem $\overline{3}$.2. *Two connected graphs homeomorphic as topological spaces have the same Euler characteristic.*

Two such graphs differ only by the number of vertices of valency 2, but that does not affect the Euler characteristic.

Let $G \subset \mathbb{R}^2$ be a connected planar graph. Then the connected components of $\mathbb{R}^2 \setminus G$ are called *faces* of the planar graph G. Let us denote by V_G, E_G, F_G the number of vertices, edges, faces of G, respectively (we omit the subscript G if it is clear from the context). We define the *Euler characteristic of the planar graph* $G \subset \mathbb{R}^2$ by setting

$$\boxed{\chi(G) := V_G - E_G + F_G}.$$

Theorem $\overline{3}$.3. *The Euler characteristic of any connected planar graph G is equal to 2:*

$$\boxed{G \subset \mathbb{R}^2 \implies \chi(G) = 2.}$$

The proof is the object of Exercise $\overline{3}$.13 (which relies on the next theorem).

Theorem $\overline{3}$.4 (Polygonal Jordan Theorem)**.** *Let C be a closed non-self-intersecting broken line (with a finite number of segments) on $\mathbb{R}^2$. Prove that $\mathbb{R}^2 \setminus C$ consists of two connected components and the boundary of each component is C.*

$\overline{3}$.4. Exercises

$\overline{3}$.1. Is it possible to build direct roads between 53 towns so that any town is connected exactly with 3 other towns?

$\overline{3}$.2. Suppose the valencies of all the vertices of a connected graph G are even. Then there exists a path that traverses each edge of G exactly once.

$\overline{3}$.3. Prove that any connected planar graph (without loops and double edges) has a vertex of degree not greater than 5.

$\overline{3}$.4. Prove that one can color the vertices of any planar graph (without loops) using five colors so that the ends of any edge have different colors.

$\overline{3}$.5. Let K_n be the graph consisting of n vertices pairwise joined by edges. Let $K_{n,m}$ be the graph consisting of $n+m$ vertices divided into two parts (n vertices in one part and m vertices in the other), the edges of $K_{n,m}$ joining each pair of vertices from different parts.

$\overline{3}$.6. Prove that the graphs $K_{3,3}$ and K_5 are not planar.

$\overline{3}$.7. (a) Let G be a planar graph such that any face of G is bounded by an even number of edges. Prove that one can color the vertices of G using two colors so that the ends of any edge have different colors.

(b) Let γ be a smooth closed curve with transversal self-intersections. Prove that γ divides the plane into domains so that one can color those domains using two colors (two domains with a common edge must be of different colors).

$\overline{3}$.8. Let a, b, c, d be points of a closed non-self-intersecting broken line C (in the plane) ordered as indicated. Suppose that points a and c are joined by a broken line L_1, points b and d are joined by a broken line L_2 and both broken lines belong to the same connected component defined by C. Prove that L_1 and L_2 have a common point.

$\overline{3}$.9. Let G be a polygonal planar graph consisting of s connected components each of which is not an isolated vertex. Let G have v vertices and e edges. Using the polygonal Jordan theorem and induction, prove that for any embedding of G in the plane the number of faces f is equal to $f=1+s-v+e$.

$\overline{3}$.10. (a) Suppose G is a planar graph without isolated vertices, v_i is the number of its vertices of degree i, f_i is the number of faces with i edges. Prove that $\sum_i(4-i)v_i+\sum_j(4-j)f_j=4(1+s)\geqslant 8$, where s is the number of connected components of G.

(b) Prove that if all faces are quadrilaterals, then $3v_1+2v_2+v_3\geqslant 8$.

(c) Prove that if the boundary of any face is a cycle containing no less than n edges, then $e\leqslant n(v-2)/(n-2)$.

$\overline{\mathbf{3}}$.11. Find and deduce the Euler Formula for convex polyhedra from the Euler formula for planar graphs. (The Euler Formula for convex polyhedra is a relation between numbers of vertices, edges and faces.)

$\overline{\mathbf{3}}$.12. With the help of Exercise $\overline{3}$.10 (c), give another proof of the nonplanarity of the graphs K_5 and $K_{3,3}$.

$\overline{\mathbf{3}}$.13. Prove Theorem $\overline{3}$.3.

Lecture 4
Examples of surfaces

In this lecture, we will study several important examples of surfaces (closed surfaces, as well as surfaces with holes) presented in different ways. We will prove that the different presentations of the same surface are indeed homeomorphic and specify their simplicial and cell space structure.

4.1. The disc $\mathbb{D}^2$

The standard two-dimensional *disk* (or 2-*disk*) is defined as

$$\mathbb{D}^2 := \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leqslant 1\}.$$

Other presentation of the 2-disk (all homeomorphic to $\mathbb{D}^2$) are: the sphere with one hole (SH), the square (Sq), the lateral surface of the cone (LC), the ellipse, the rectangle, the triangle, the hexagon, etc. (see Figure 4.1).

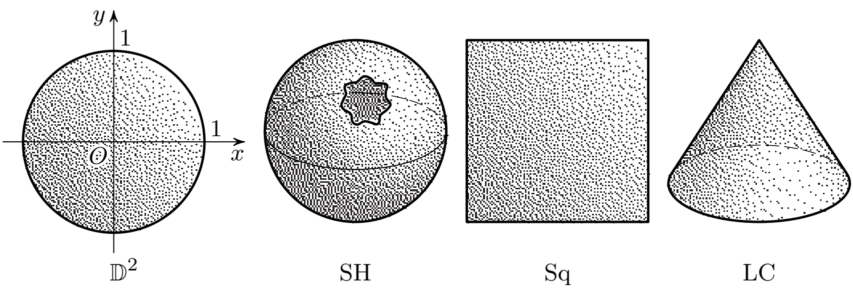


FIGURE 4.1. Different presentations of the disk

The simplest cell space structure of the 2-disk consists of one 0-cell, one 1-cell, and one 2-cell, but of course other cell space structures are possible.

It is easy to prove that the different presentations of the disk listed above are homeomorphic. For example, a homeomorphism of $\mathbb{D}^2$ onto the square Sq is obtained by centrally projecting concentric circles filling $\mathbb{D}^2$ onto the corresponding circumscribed concentric square boundaries. More precisely, we define $h\colon \mathbb{D}^2 \to \mathrm{Sq}$ as follows: let a point $P \in \mathbb{D}^2$ be given; denote by C_P the circle centered at the center O of the disk and passing through P; denote by D_P the boundary of the square with sides parallel to the sides of Sq circumscribed to C_P; then $h(P)$ is defined as the intersection of the ray $[OP\rangle$ with D_P. It is easy to see that h is a homeomorphism, so that the disk $\mathbb{D}^2$ and the square Sq are indeed homeomorphic.

Describing the other homeomorphisms of $\mathbb{D}^2$ (onto the sphere with one round hole (SH), the lateral surface of the cone (LC), the ellipse, the rectangle) is the object of Exercise 4.1.

4.2. The sphere $\mathbb{S}^2$

The standard two-dimensional *sphere* (or 2-*sphere*) is defined as

$$\mathbb{S}^2 := \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\}.$$

Other presentations of the 2-sphere (all homeomorphic to $\mathbb{S}^2$) include: the boundary of the cube or the tetrahedron, the disk with boundary identified to one point $\mathbb{D}^2/\partial\mathbb{D}^2$, the suspension over the circle $\Sigma(\mathbb{S}^2)$, the join of the circle and the 0-sphere (i.e., a pair of points) $\mathbb{S}^2 * \mathbb{S}^0$, the boundary of any closed convex body, the configuration space of the 3-dimensional pendulum (the line segment in $\mathbb{R}^3$ with one extremity fixed), etc.

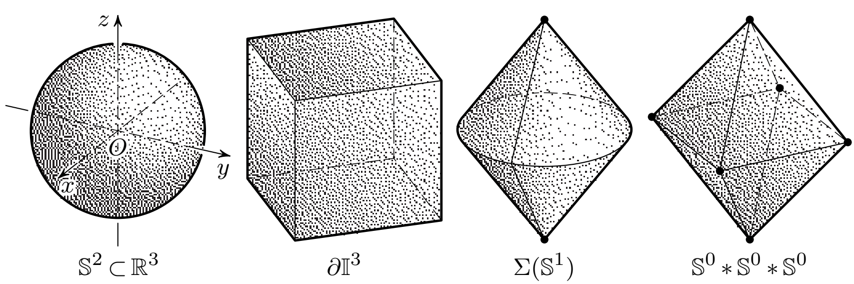


FIGURE 4.2. Different presentations of the sphere

The simplest cell space structure of the 2-sphere consists of one 0-cell and one 2-cell.

Homeomorphisms between the various presentation of $\mathbb{S}^2$ listed above are easy to construct (central projection is the main instrument here; see Exercise 4.1).

4.3. The Möbius band Mb

The Möbius band (or Möbius strip) Mb is usually modeled by a long rectangular strip of paper with the two short sides identified ("glued together") after a half twist (Figure 4.3). Formally it can be defined as the square with two opposite sides identified $\mathrm{Sq}/_{\sim}$ via the central symmetry $\sim$. A beautiful embedding of the Möbius strip $\mathrm{Mb} \hookrightarrow \mathbb{R}^3$ can be observed as a trefoil knot spanned by a soap film; the same embedded surface can be obtained by giving a long strip of paper three half-twists and then identifying the short sides. An even more complicated embedding of the Möbius strip in $\mathbb{R}^3$ is obtained by giving a long strip of paper a large odd number of half-twists and then identifying the short sides.

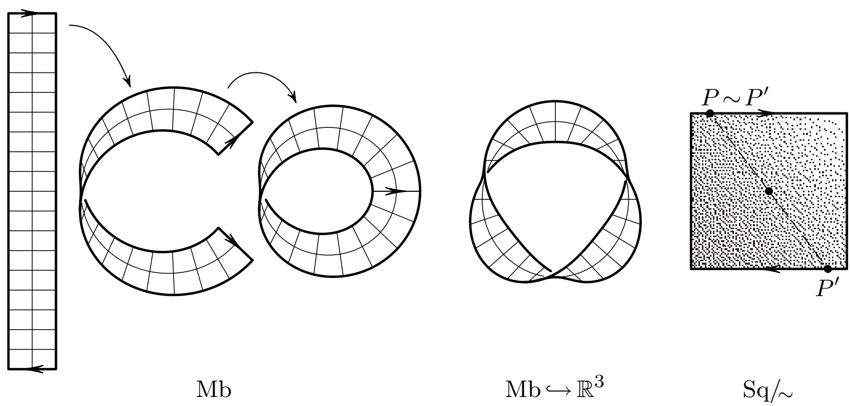


Figure 4.3. Different presentations of the Möbius strip

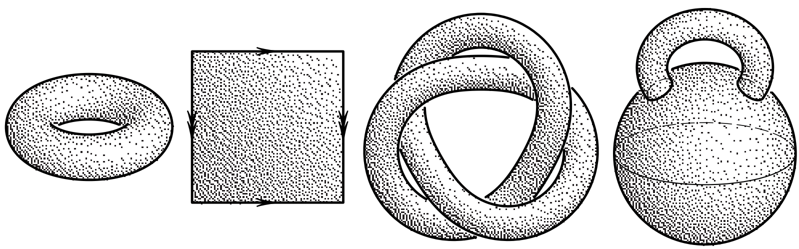
Everyone knows that the Möbius strip is "one-sided" (it cannot be painted in two colors) and is "nonorientable". (The definition of "nonorientable surface" will be given below.) If you have never done this before, try to guess what happens to the Möbius strip if you cut it along its midline. Check the validity of your guess by using scissors on a paper model.

4.4. The torus $\mathbb{T}^2$

Topologically, the (2-dimensional) torus $\mathbb{T}^2$ is defined as the Cartesian product of two circles. Geometrically, it can be presented as the set of points $(x, y, z) \in \mathbb{R}^3$ satisfying the equation

$$\left(x^2+y^2+z^2+R^2+r^2\right)^2-4R^2\left(x^2+y^2\right)=0.$$

The torus can also be presented as the square with opposite sides identified $\mathrm{Sq}/_\sim$ (the identifications are shown by the arrows in Figure 4.4), as a surface embedded (in different ways) in 3-space $\mathbb{T}^2 \hookrightarrow \mathbb{R}^3$, as a "sphere with one handle" M_1^2, as an annulus with boundary circles (oriented in the same direction) identified, as the configuration space of the double pendulum with arms $L > l$, as the plane $\mathbb{R}^2$ modulo the periodic equivalence $(x, y) \sim (x+1, y+1)$, etc.



$\mathbb{T}^2 \subset \mathbb{R}^3$ $\mathrm{Sq}/_\sim$ $\mathbb{T}^2 \hookrightarrow \mathbb{R}^3$ M_1^2

FIGURE 4.4. Different presentations of the torus

4.5. The projective plane $\mathbb{R}P^2$

The *projective plane* $\mathbb{R}P^2$ is defined as the set of straight lines l in $\mathbb{R}^3$ passing through the origin, with the natural topology (its base consists all open cones around all elements $l \in \mathbb{R}P^2$). The notion of straight line is naturally defined in $\mathbb{R}^2$: a "line" is a (Euclidean) plane P passing through the origin, its "points" are all the (Euclidean) lines l passing through the origin and contained in P.

Each element $l \in \mathbb{R}P^2$ may be specified by its homogeneous coordinates, i.e., the three coordinates of any point (of $\mathbb{R}^3$) on the (Euclidean) line l considered up to a common factor λ, so that $(x : y : z)$ and $(\lambda x : \lambda y : \lambda z)$, $\lambda \neq 0$, specify the same point of $\mathbb{R}P^2$.

Other presentations (Figure 4.5) of $\mathbb{R}P^2$ are: the disk $\mathbb{D}^2$ with diametrically opposed boundary points identified $\mathbb{D}^2/_\sim$, the sphere $\mathbb{S}^2$ with all

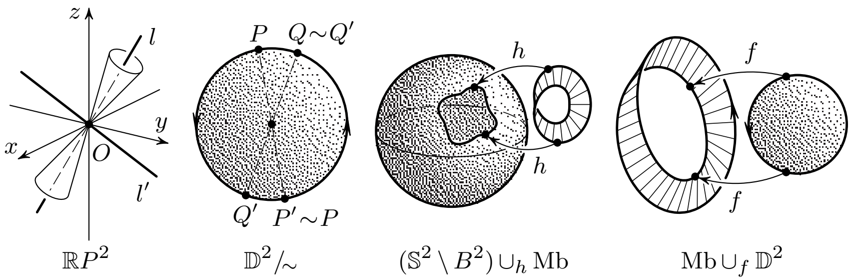


FIGURE 4.5. Different presentations of the projective plane

pairs of points symmetric with respect to the origin identified $\mathbb{S}^2/\mathrm{Ant}$, the sphere with a hole with a Möbius strip attached to it along the boundary $(\mathbb{S}^2 \setminus B^2) \cup_h \mathrm{Mb}$, the Möbius band with a disk glued to it along the boundary $\mathrm{Mb} \cup_f \mathbb{D}^2$, the square with centrally symmetric boundary points identified, the configuration space of a rectilinear rod rotating in $\mathbb{R}^3$ about a fixed hinge at its midpoint. The proof that all these presentations are homeomorphic is pleasant and straightforward (see Exercise 4.2).

The simplest cell space structure on $\mathbb{R}P^2$ consists of one cell in each dimension 0, 1, 2 and can be easily seen on the disk model. Note that the boundary of the 2-cell wraps around the 1-cell twice.

4.6. The Klein bottle Kl

The *Klein bottle* can be defined as the square with opposite sides identified as shown by the arrows in Figure 4.6. The Klein bottle cannot be embedded into $\mathbb{R}^3$ (see Exercise 4.12), and so we cannot draw a realistic picture of it. The Klein bottle Kl is usually pictured as in Figure 4.6, but

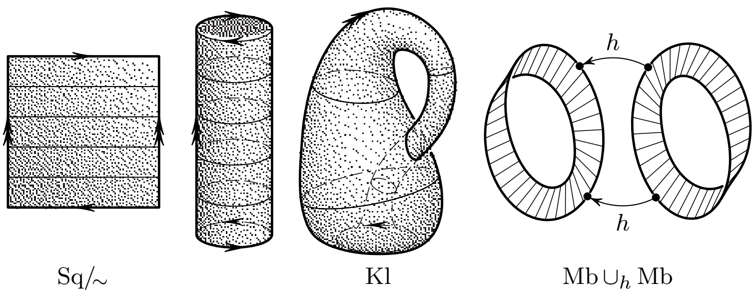


FIGURE 4.6. Different presentations of the Klein bottle

that picture is not correct: the "surface" in the figure has a self-intersection (a little circle), so it is not homeomorphic to Kl.

Here are some other presentations of the Klein bottle: two Möbius strips identified along their boundary circles $\mathrm{Mb} \cup_h \mathrm{Mb}$, two projective planes with holes with the boundaries of the holes identified, etc.

4.7. The disk with two holes ("pants")

This surface is obtained from the disk $\mathbb{D}^2$ by removing two small open disks from $\mathbb{D}^2$; it is called *pants* by topologists and denoted $\mathbb{P}$. It plays an important technical role in low-dimensional topology, in particular in the next lecture.

It is possible to construct a torus (the sphere with one handle) from two copies of pants (glue the boundaries of the four "legs" together and then close up the two "waists" by gluing disks to them). In a similar way, we can construct a sphere with 2, 3, 4, ... handles.

Different ways of presenting the disk with two holes are shown in Figure 4.7.

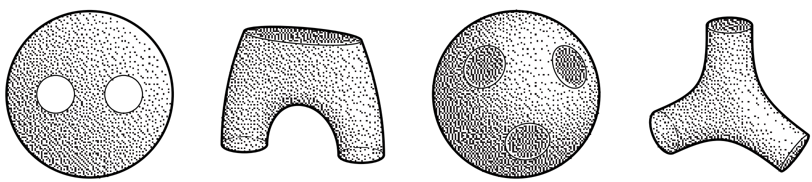


FIGURE 4.7. Different presentations of the disk with two holes

4.8. Triangulated surfaces

The surfaces (with or without holes) described above can easily be triangulated, i.e., supplied with the structure of a (two-dimensional) simplicial space. Simple examples of triangulations are shown in Figure 4.8. For triangulated surfaces the holes are usually chosen as the insides of 2-simplices, so that the boundaries of the holes will be triangles consisting of three 1-simplices. Any 1-simplex which is not on part of a boundary is the common side of two triangles (2-simplices).

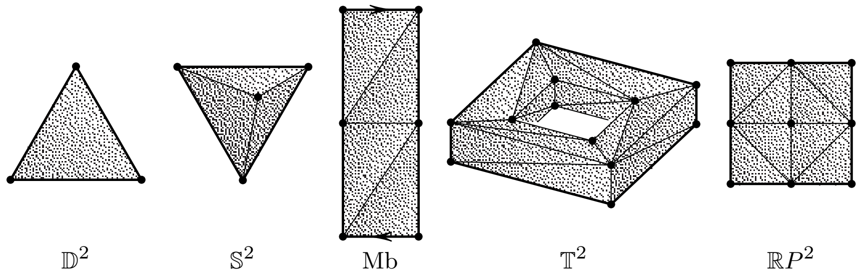


FIGURE 4.8. Triangulations of the disk, the sphere, the Möbius strip, the torus, and the projective plane (left to right).

4.9. Orientable surfaces

A triangulated surface is called orientable if all its 2-simplexes can be "oriented coherently". We do not explain what this means because, for topological surfaces, orientability can be defined in a simpler way: a surface M is called *orientable* if it does not contain a Möbius strip, and is called *nonorientable* otherwise.

It is easy to prove that *the Möbius strip, the projective plane, and the Klein bottle are nonorientable.* It is intuitively clear that *the disk, the sphere, the torus, the pants are orientable*, but this is not easy to prove. (We will come back to this question in the next lecture).

4.10. Euler characteristic

Let M be a triangulated surface, for example one of the triangulated surfaces described in Section 4.8. Then the *Euler characteristic* of M, denoted by $\chi(M)$, is defined as

$$\boxed{\chi(M) := V - E + F},$$

where V is the number of vertices (0-simplices), E is the number of edges (1-simplices), and F is the number of faces (2-simplices) in the triangulation of the surface M.

It will be shown in the next lecture that *the Euler characteristic does not depend on the choice of triangulation, i.e., $\chi(M)$ is a topological invariant*:

$$M \approx M' \implies \chi(M) = \chi(M').$$

Theorem 4.1. *The Euler characteristics of the disk, the sphere, the torus, the pants, the Möbius strip, the projective plane, and the Klein bottle are respectively equal to* $1, 2, 0, -1, 0, 1, 0$.

Since we know that $\chi(M)$ does not depend on the choice of triangulation, to prove the theorem it suffices to compute $\chi(M)$ (using its definition) for the triangulated surfaces described in Section 4.8.

4.11. Connected sum

Given two surfaces M_1 and M_2, their *connected sum* $M_1 \# M_2$ is obtained by removing little open disks from each and gluing them together along a homeomorphism of the little boundary circles of the removed disks. In the case when M_1 and M_2 are triangulated, it is more convenient to remove the interior of a 2-simplex in each surface and glue them together along a piecewise linear homeomorphism of the boundaries of the removed simplices.

For $M_1 \# M_2$ to be well-defined, we should prove that the connected sum does not depend on the position of the removed open disks and on the choice of the attaching homeomorphism. This can be done by a technical argument that we omit.

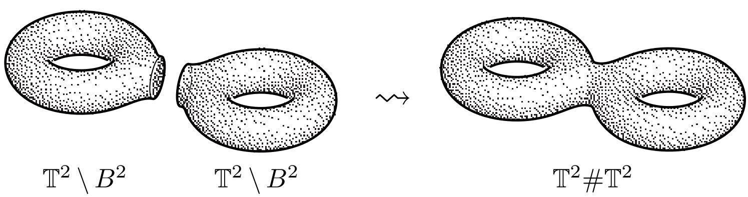


FIGURE 4.9. Connected sum of two tori

Knowing the Euler characteristics of two given surfaces M_1 and M_2, it is easy to compute the Euler characteristic of their connected sum $M_1 \# M_2$: two faces (2-simplices) have disappeared, three edges (1-simplices) have been identified with three other edges, three vertices (0-simplices) have been identified with three other vertices, so that the Euler characteristic of the connected sum is 2 less than the sum of the Euler characteristics of the summands. We have proved the following theorem:

Theorem 4.2. $\chi(M_1 \# M_2) = \chi(M_1) + \chi(M_2) - 2$.

4.12. Exercises

4.1. Show that the surfaces in Figure 4.1, the surfaces in Figure 4.2, the surfaces in Figure 4.3, are homeomorphic.

4.2. Prove that the projective plane is (a) the Möbius strip with a disk attached; (b) the sphere S^2 with antipodal points identified; (c) the disk D^2 with diametrically opposed points identified.

4.3. Prove that the Klein bottle is (a) the double of the Möbius strip; (b) the sphere with two holes with two Möbius strips attached; (c) the connected sum of two projective planes.

4.4. (a) Consider the topological space of straight lines in the plane. Prove that this space is homeomorphic to the Möbius band without boundary.

(b) Consider the topological space of *oriented* straight lines in the plane. Prove that this space is homeomorphic to the cylinder without boundary.

4.5. Show that a punctured tube from a bicycle tire can be turned inside out. (More precisely, this would be possible if the rubber from which the tube is made were elastic enough.)

4.6. (a) *Polygonal Schoenflies Theorem.* A closed polygonal line in the plane bounds a domain whose closure is the disk D^2.

(b) *Polygonal Annulus Theorem.* Two closed polygonal lines in the plane, one of which encloses the other, bound a domain whose closure is the annulus $S^1 \times [0, 1]$.

4.7. (a) The two surfaces with holes obtained from the same closed triangulated connected surfaces by removing different open 2-simplices from it are homeomorphic. (b) Show that the connected sum of surfaces is well defined.

4.8. Prove that $\mathbb{T}^2 \# \mathbb{R}P^2 \approx 3\mathbb{R}P^2$.

4.9. (a) Prove that $\mathrm{Kl}\#\mathrm{Kl}$ is homeomorphic to the Klein bottle with one handle attached. (b) Prove that $\mathbb{R}P^2\#\mathrm{Kl}$ is homeomorphic to the projective plane with one handle attached.

4.10. Prove that if a surface M_1 is nonorientable, then for any surface M_2 the surface $M_1 \# M_2$ is nonorientable.

4.11. How many different surfaces is it possible to glue (by identifying sides) starting with (a) a square; (b) a hexagon; (c) an octagon.

4.12. Prove that the Klein bottle cannot be embedded in $\mathbb{R}^3$. (*Hint*: you can use the fact that the graph $K_{3,3}$ cannot be embedded in $\mathbb{R}^2$).

Lecture 5
Classification of surfaces

In this lecture, we will present the topological classification of surfaces. This will be done by a combinatorial argument imitating Morse theory and will make use of the Euler characteristic.

5.1. Main definitions

In this course, by a *surface* we mean a connected compact topological space M such that that any point $x \in M$ possesses an open neighborhood $U \ni x$ whose closure is a 2-dimensional disk. By a *surface-with-holes* ("поверхность с краем" in Russian) we mean a a connected compact topological space M such that any point $x \in M$ possesses either an open neighborhood $U \ni x$ whose closure is a 2-dimensional disk, or a whose closure is the open half disk

$$C = \{(x, y) \in \mathbb{R}^2 \colon y \geqslant 0, x^2 + y^2 < 1\}.$$

(A synonym of "surface" is "two-dimensional compact connected manifold", but we will use the shorter term.) In the previous lecture, we presented several examples of surfaces and surfaces-with-holes.

It easily follows from the definitions that the set of all points of a surface-with-holes that have half-disk neighborhoods is a finite family of topological circles. We call each such circle the *boundary* of a hole. For example, the Möbius strip has one hole, pants have three holes.

5.2. Triangulating surfaces

In the previous lecture, we gave examples of triangulated surfaces (see Fig. 4.8). Actually, it can be proved that *any* surface (or any surface-with holes) can be triangulated, but the known proofs are difficult, rather ugly, and based on the Jordan Curve Theorem (whose known proofs are also difficult). So we will accept this as a fact without proof.

Fact 5.1. *Any surface and any surface-with-holes can be triangulated.*

To state the next fact about triangulated surfaces, we need some definitions. Recall that a (continuous) map $f\colon M \to N$ of triangulated surfaces is called *simplicial* if it sends each simplex of M onto a simplex of N (not necessarily of the same dimension) linearly. Any bijective simplicial map map $f\colon M \to N$ is said to be an *isomorphism*, and then M and N are called *isomorphic*.

Suppose M is a triangulated surface, σ^2 is a face of M and w is an interior point of σ^2. Then the new triangulation of M obtained by joining w to the three vertices of σ^2 is called a *face subdivision* of M at σ (Fig. 5.1 (a)); the *barycentric subdivision* of a 2-simplex is shown in Fig. 5.1 (c); the *barycentric subdivision of M* is obtained by barycentrically subdividing all its 2-simplices. If σ^1 is an edge (1-simplex) of M, then the *edge subdivision* of M at σ^2 is shown on Fig. 5.1 (b). If a triangulated surface M' is obtained from M by subdividing some simplices of M in some way, we say that M' is a *subdivision of M*.

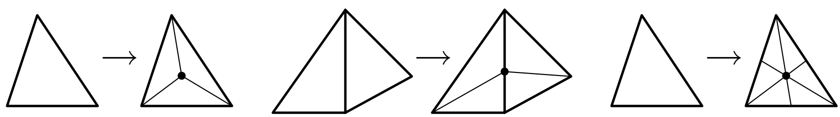


FIGURE 5.1. Face, edge, and barycentric subdivisions

A map $f\colon M \to N$ is called a PL-*map* if there exist subdivisions of M', N' of M, N such that f is a simplicial map of M' to N'. A bijective PL-map $f\colon M \to N$ is said to be a PL-*equivalence*, and then M and N are called PL-*equivalent*. The following statement, known as the *Hauptvermutung for surfaces*, will be stated without proof.

Fact 5.2. *Two surfaces are homeomorphic if and only if they are PL-equivalent. Homeomorphic triangulated surfaces have isomorphic triangulations.*

If x, y are vertices of M, then the *star* of x, $\mathrm{St}(x)$, is defined as the union of all simplices for which x is a vertex, and the *link* of y, $\mathrm{Lk}(y)$, is the union of all 1-simplices opposite to the vertex y of the 2-simplices forming $\mathrm{St}(x)$. It is easy to show that $\mathrm{St}(x)$ is, topologically, a 2-disk, and $\mathrm{Lk}(y)$, a circle (see Figure 5.2).

In the previous lecture, orientable surfaces were defined as surfaces not containing a Möbius strip. Now we give another (equivalent) definition of orientability for triangulated surfaces. A simplex $\sigma^2 = [0, 1, 2]$ is called

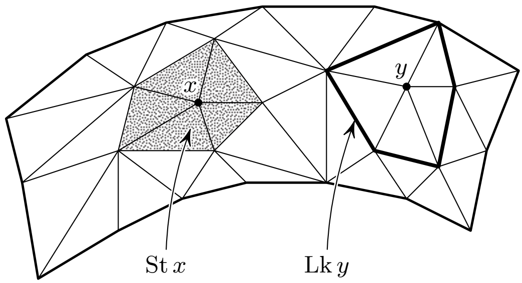


FIGURE 5.2. Star and link of points on a surface

oriented if a cyclic order of its vertices is chosen. Adjacent oriented simplices are *coherently oriented* if their common edge acquires opposite orientations induced by the two oriented simplices. Thus if the two simplices $\sigma_1^2 = [0, 1, 2]$ and $\sigma_2^2 = [0, 1, 3]$ are coherently oriented if the cyclic orders chosen in the two simplices are $(0, 1, 2)$ and $(1, 0, 3)$, respectively. A triangulated surface is called *orientable* if all its 2-simplices can be coherently oriented.

It is not hard to prove that *a surface is orientable if and only if it does not contain a Möbius strip.*

5.3. Classification of orientable surfaces

The main result of this section is the following theorem.

Theorem 5.3 (Classification of orientable surfaces)**.** *Any orientable surface is homeomorphic to one of the surfaces in the following list*

$$\mathbb{S}^2,\ \mathbb{S}^1 \times \mathbb{S}^1\ (\textit{torus}),\ (\mathbb{S}^1 \times \mathbb{S}^1)\#(\mathbb{S}^1 \times \mathbb{S}^1)\ (\textit{sphere with } 2 \textit{ handles}),\ \dots$$
$$\dots,\ (\mathbb{S}^1 \times \mathbb{S}^1)\#(\mathbb{S}^1 \times \mathbb{S}^1)\#\dots\#(\mathbb{S}^1 \times \mathbb{S}^1)\ (\textit{sphere with } k \textit{ handles}),\ \dots$$

Any two (*different*) *surfaces in the list are not homeomorphic.*

Proof. In view of Fact 5.1, we can assume that M is triangulated and take the double baricentric subdivision M'' of M. In this triangulation, the star of a vertex of M'' is called a *cap*, the union of all faces of M'' intersecting an edge of M but not contained in the caps is called a *strip*, and the connected components of the union of the remaining faces of M'' are called *patches*.

Consider the union of all the edges of M; this union is a graph (denoted G). Let G_0 be a maximal tree of G. Denote by M_0 the union of

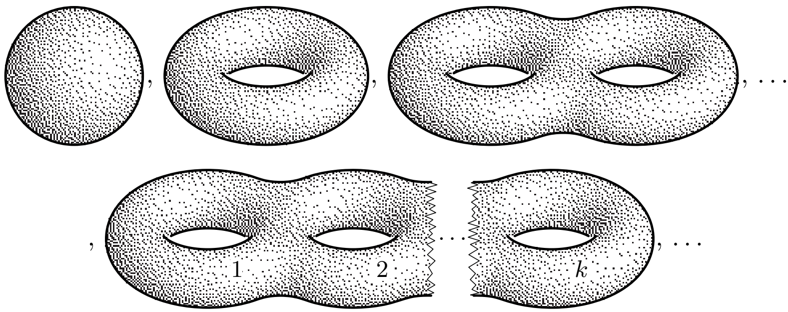


Figure 5.3. The orientable surfaces

all caps and strips surrounding G_0. Clearly M_0 is homeomorphic to the 2-disk (why?). If we successively add the strips and patches from $M - M_0$ to M_0, obtaining an increasing sequence

$$M_0 \subset M_1 \subset M_2 \subset \ldots \subset M_p = M,$$

we shall recover M.

Let us see what happens when we go from M_0 to M_1.

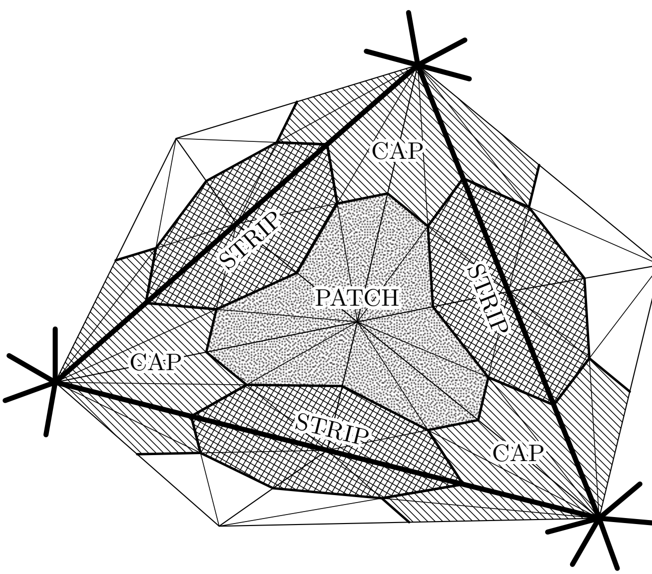


Figure 5.4. Caps, strips, and patches

If there are no strips left[1], then there must be a patch (topologically, a disk), which is attached along its boundary to the boundary circle Σ_0 of M_0; the result is a 2-sphere and the theorem is proved.

Suppose there are strips left. At least one of them, say S, is attached along one end to Σ_0 (because M is connected) and its other end is also attached to Σ_0 (otherwise S would have been part of M_0). Denote by K_0 the closed *collar neighborhood* of Σ_0 in M_0 (i.e., the union of all simplices having at least one vertex on Σ_0). The collar K_0 is homoeomorphic to the annulus (and not to the Möbius strip) because M is orientable. Attaching S to M_0 is the same as attaching another copy of $K_0 \cup S$ to M_0 along Σ_0'. But $K \cup S$ is homeomorphic to the disk with two holes (what we have called "pants"), because S has to be attached in the orientable way in view of the orientability of M (for that reason the twisting of the strip shown in Figure 5.5 (a) cannot occur). Thus M_1 is obtained from M_0 by attaching the pants $K \cup S$ by the waist (which is Σ_0'), and M_1 has two boundary circles (Figure 5.5 (b)).

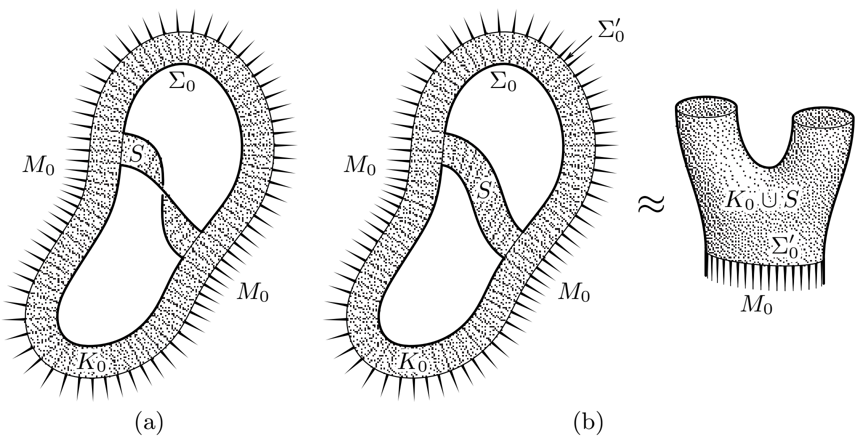


FIGURE 5.5. Adding pants along the legs

Now let us see what happens when we pass from M_1 to M_2.

If there are no strips left, there are two patches that must be attached to the two boundary circles of M_1, and we get the 2-sphere again.

[1] Actually, this case cannot occur, but it is more complicated to prove this than to prove that the theorem holds in this case.

Suppose there are patches left. Pick one, say S, which is attached at one end to one of the boundary circles, say Σ_1 of M_1. Two cases are possible: either

(i) the second end of S is attached to Σ_2, or
(ii) the second end of S is attached to Σ_1.

Consider the first case. Take collar neighborhoods K_1 and K_2 of Σ_1 and Σ_2; both are homoeomorphic to the annulus (because M is orientable). Attaching S to M_1 is the same as attaching another copy of $K_1 \cup K_2 \cup S$ to M_1 along the two circles Σ_1' and Σ_2' (because the copy of $K_1 \cup K_2$ can be homeomorphically pushed into the collars K_1 and K_2). But $K_1 \cup K_2 \cup S$ is obviously homeomorphic to the disk with two holes. Thus, in the case considered, M_2 is obtained from M_1 by attaching pants to M_1 along the legs, thus decreasing the number of boundary circles by one.

The second case is quite similar to adding a strip to M_0 (see above), and results in attaching pants to M_1 along the waist, increasing the number of boundary circles by one.

What happens when we add a strip at the i-th step? As we have seen above, two cases are possible: either the number of boundary circles of M_{i-1} increases by one or it decreases by one. We have seen that in the first case "inverted pants" are attached to M_{i-1} and in the second case "upright pants" are added to M_{i-1}.

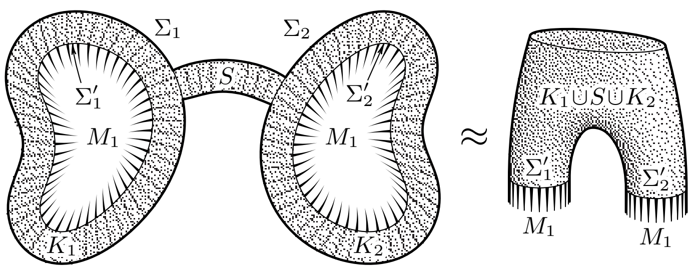


FIGURE 5.6. Adding pants along the waist

After we have added all the strips, what will happen when we add the patches? The addition of each patch will "close" a pair of pants either at the "legs" or at the "waist". As the result, we obtain a surface. Let us prove that this surface is a sphere with m handles, $m \geqslant 0$.

We prove this by induction on the number k of attached pants. The base of induction ($k = 0$) was established above. Assume that by attaching

$k-1$ pants by the waist and by the legs and patching up (attaching disks to the free boundaries) we always obtain a sphere with some number ($\geqslant 0$) of handles. Let us prove that this will be true for k pants. We will consider two cases.

Case 1: The last pants were attached by the waist (and then the legs were patched up). Removing the pants (together with the two patches) from our surface M and patching up the waist W, we obtain a surface M_1 constructed from $k-1$ pants. By the induction hypothesis, M_1 is a sphere with $m_1 \geqslant 0$ handles. But M is obtained from M_1 by removing the patch of W and attaching pants by the waist and patching up. But then M is obviously a sphere with the same number (m_1) of handles.

Case 2: The last pants were attached by the legs (and then the waist was patched up). Removing the pants (together with the two patches) from our surface M and patching up the waist W, we obtain a surface M_1 constructed from $k-1$ pants. By the induction hypothesis, M_1 is a sphere with $m_2 \geqslant 0$ handles. But M is obtained from M_1 by removing the patch of W and attaching pants by the waist and patching up. But then M is obviously a sphere with (m_1+1) handles.

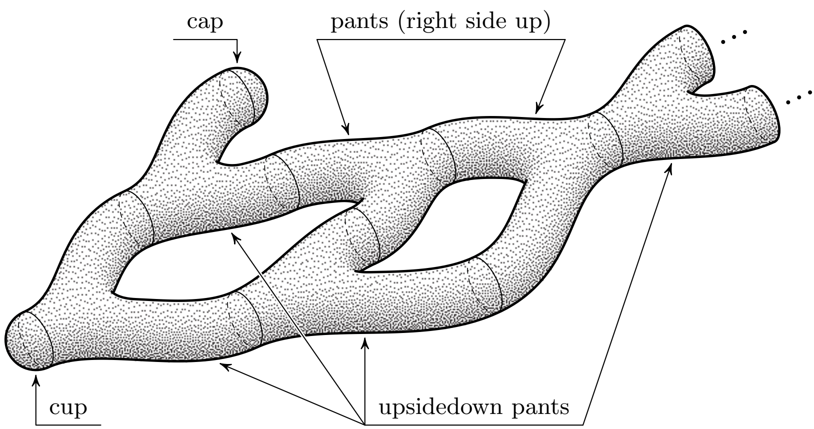


Figure 5.7. Constructing an orientable surface

The first part of the theorem is proved.

To prove the second part, it suffices to show that

(1) *homeomorphic surfaces have the same Euler characteristic*;

(2) *all the surfaces in the list have different Euler characteristics* (*namely* $2, 0, -2, -4, \dots,$ *respectively*).

The first statement follows from Fact 5.2. Indeed, if two surfaces are homeomorphic, then they have isomorphic subdivisions. It is easy to verify that *the Euler characteristic does not change under subdivision.* To do that, it suffices to check that the Euler characteristic does not change under face, edge, barycentric subdivision, which is straightforward. This proves (1).

The second statement is proved by simple computations using the formula for the Euler characteristic of a connected sum (Theorem 4.2).

The theorem is proved. □

The *genus* g of an orientable surface can be defined as the number of its handles and can be expressed in terms of the Euler characteristic in the following way:

$$g(M) = \frac{1}{2}\Big(2 - \chi(M)\Big).$$

In fact, this has already been established in the above computation of the Euler characteristic of orientable surfaces.

5.4. Classifying nonorientable surfaces and surfaces-with-holes

Theorem 5.4. *Any nonorientable surface is contained in the following list:*

$$\mathbb{R}P^2,\ \mathbb{R}P^2\#\mathbb{R}P^2,\ \dots,\ \mathbb{R}P^2\#\mathbb{R}P^2\#\dots\#\mathbb{R}P^2\ (g\ \textit{summands}),\ \dots$$

Two different surfaces in the list are not homeomorphic.

We omit the proof (similar to that of Theorem 5.3, but more complicated).

The nonnegative integer $g = 1 - \chi(N)$ is called the *genus* of the nonorientable surface N. Thus the genus of the Klein bottle is 1, i.e., it is equal to its “number of handles” in a natural sense. This is also true for the the other nonorientable surfaces (see Exercise 5.11).

We leave the statement of the classification theorem of all surfaces-with-holes to the reader.

5.5. Exercises

5.1. Prove that $\chi(m\mathbb{T}^2) = 2 - 2m$ and $\chi(n\mathbb{R}P^2) = 2 - n$. (Here the notation nM stands for the connected sum of n copies of M.)

5.2. Prove that an orientable surface is not homeomorphic to a nonorientable surface.

5.3. (a) Prove that any graph has a maximal subtree. (b) Prove that a simplicial neighborhood of a tree in a surface is homeomorphic to the disk.

5.4. Find the Euler characteristic of the Klein bottle.

5.5. Consider the quotient space $(S^1 \times S^1)/\big((x, y) \sim (y, x)\big)$. This space is a surface. Which one?

5.6. Show that the standard circle can be spanned by a Möbius band, i.e., the Möbius band can be homeomorphically deformed in 3-space so that its boundary becomes a circle lying in some plane.

5.7. Prove that the boundary of $\mathrm{Mb} \times [0, 1]$ is the Klein bottle.

5.8. Prove that on the sphere with g handles, the maximal number of nonintersecting closed curves not dividing this surface is equal to g.

5.9. Can $K_{3,3}$ be embedded (a) in the sphere; (b) in the torus; (c) in the Klein bottle; (d) in the Möbius strip?

5.10. Prove that the Klein bottle cannot be embedded in $\mathbb{R}^3$.

5.11. Prove that $\mathbb{T}\#\mathbb{R}P^2$ is homeomorphic to $3\mathbb{R}P^2$ and more generally $m\mathbb{T}^2\# \approx (m+2)\mathbb{R}P^2$.

Lecture 6

Homotopy

The notions of homotopy and homotopy equivalence are quite fundamental in topology. Homotopy equivalence of topological spaces is a weaker equivalence relation than homeomorphism, and *homotopy theory* studies topological spaces up to this relation (and maps up to homotopy). This theory constitutes the main body of *algebraic topology*, but we only consider a few of its basic notions here. One of these notions is the Euler characteristic, which is also a homotopy invariant.

6.1. Homotopic maps

Two maps $f, g\colon X \to Y$ are called *homotopic* (notation $f \simeq g$) if they can be joined by a *homotopy*, i.e., by a map $F\colon X \times [0, 1] \to Y$ such that $F(x, 0) \equiv f(x)$ and $F(x, 1) \equiv g(x)$ (here $\equiv$ means for all $x \in X$). If we change the notation from $F(x, t)$ to $F_t(x)$, we can restate the previous definition by saying that there exists a family $\{F_t(x)\}$ of maps, parametrized by $t \in [0, 1]$, continiously changing from $f \equiv F_0$ to $g \equiv F_1$.

It is easy to prove that

$f \simeq f$ for any $f\colon X \to Y$ (*reflexivity*);

$f \simeq g \Longrightarrow f \simeq g$ for all $f, g\colon X \to Y$ (*symmetry*);

$f \simeq g$ and $g \simeq h \Longrightarrow f \simeq h$ for all $f, g, h\colon X \to Y$ (*transitivity*).

For example, to prove transitivity, we obtain a homotopy joining f and h by setting

$$F(x, t) = \begin{cases} F_1(x, 2t) & \text{for } 0 \leqslant t \leqslant 1/2, \\ F(x, t) = F_2(x, 2t-1) & \text{for } 1/2 \leqslant t \leqslant 1, \end{cases}$$

where F_1, F_2 are homotopies joining f and g, g and h, respectively.

Thus the homotopy of maps is an equivalence relation, so that the set $\mathrm{Map}(X, Y)$ of all (continuous) maps of X to Y splits into equivalence classes, called *homotopy classes*; the set of these equivalence classes is denoted $[X, Y]$.

6.2. Homotopy equivalence of spaces

Two spaces X and Y are called *homotopy equivalent* if there exist two maps $f\colon X \to Y$, $g\colon Y \to X$ (called *homotopy equivalences*) such that $f \circ g \simeq \mathrm{id}_Y$ and $g \circ f \simeq \mathrm{id}_X$.

Obviously, *homeomorphic spaces are homotopy equivalent* (the homotopy equivalences are provided by any homeomorphism and its inverse). The converse statement is not true: for example, the point is homotopy equivalent to the 2-disk, but these two spaces are not homeomorphic.

Thus homotopy equivalence is a weaker equivalence relation than homeomorphism, so that homotopy classification is rougher (and hence easier—there are less classes) than topological classification. Its importance in topology is due to the fact that most topological invariants are homotopy invariants (this is the case of the so-called fundamental group, homology groups and related invariants such as the Euler characteristic).

6.3. Degree of maps of $\mathbb{S}^1$ into itself

In this section we consider (continuous) maps $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ of the circle into itself. Examples are the maps $w_k\colon \mathbb{S}^1 \to \mathbb{S}^1$ given by the rule $e^{i\varphi} \mapsto e^{ik\varphi}$, where $\mathbb{S}^1$ is modeled by the set of unimodular complex numbers: $\mathbb{S}^1 = \{z \in \mathbb{C}\colon |z| = 1\}$. We are interested in the homotopy classification of such maps taking the basepoint $1 \in \mathbb{S}^1$ to the basepoint $1 \in \mathbb{S}^1$.

Theorem 6.1. *There is a natural bijection between homotopy classes of maps of the circle into itself and the integers given by the degree* deg *(defined below)*

$$\boxed{\deg\colon [\mathbb{S}^1, \mathbb{S}^1] \longrightarrow \mathbb{Z}.}$$

Proof. Consider the map $\exp\colon \mathbb{R} \to \mathbb{S}^1$ given by the rule $\mathbb{R} \ni \varphi \mapsto e^{i\varphi} \in \mathbb{S}^1$. The map exp is not a bijection; for example, it takes all points of the form $2k\pi$ to $1 \in \mathbb{S}^1$ (see Fig. 6.1 (a)). However, exp is a *local homeomorphism*, i.e., any point has a neighborhood U (e. g. any open interval of length less than 2π containing the point) such that the restriction $\exp|_U$ of exp to U is a homeomorphism.

Now any map $\mathbb{S}^1 \to \mathbb{S}^1$ can be regarded as a map $f\colon [0, 2\pi] \to \mathbb{S}^1$ such that $f(0) = f(2\pi) = 1 \in \mathbb{S}^1$. For any such map there exists a unique map $\widetilde{f}\colon [0, 2\pi] \to \mathbb{R}$, called the *lift* of f, such that $\exp \circ \widetilde{f} = f$ and $\widetilde{f}(0) = 0$.

Indeed, subdivide $[0, 2\pi]$ into segments $[0, a_1], [a_1, a_2], \dots, [a_m, 2\pi]$, so small that none of the images of these segments covers $\mathbb{S}^1$; then, using the fact that exp is homeomorphic on each segment, successively extend the map taking the point $0 \in [0, 2\pi]$ to the point $0 \in \mathbb{R}$ to a map $\widetilde{f}$ of the whole interval $[0, 2\pi]$ to $\mathbb{R}$. (Look at Figure 6.1 (a).)

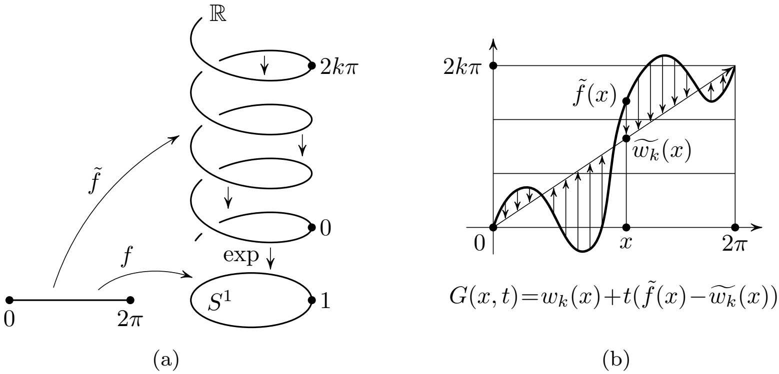


FIGURE 6.1. Liftings of the exponential map

We now define the *degree* $\deg([f])$ of any circle map $[f] \ni f\colon \mathbb{S}^1 \to \mathbb{S}^1$ as follows:

$$\boxed{\deg([f]) := \widetilde{f}(2\pi)/2\pi.}$$

To prove the theorem, we must show that:

(0) $\widetilde{f}(2\pi)$ does not depend on the choice of points $a_1, \dots, a_m$ that subdivide $[0, 2\pi]$;

(1) the assignment $\deg\colon [\mathbb{S}^1, \mathbb{S}^1] \to \mathbb{Z}$ is well defined, i.e., if f is homotopic to f', then $\deg([f]) = \deg([f'])$;

(2) the assignment deg is injective, i.e., if f is not homotopic to f', then $\deg([f]) \neq \deg([f'])$;

(3) the assignment deg is surjective, i.e., for any $k \in \mathbb{Z}$ there exists a map f such that $\deg[f] = k$.

To do this, we will need a lemma.

Lemma. *If $\widetilde{f}\colon [0, 2\pi] \to \mathbb{R}$ and $k = \widetilde{f}(2\pi)/2\pi$, then $\widetilde{f}$ is homotopic to $\widetilde{w_k}$, where $\widetilde{w_k}$ is the lift of w_k. Moreover, f will then be homotopic to w_k.*

Proof. Look at Fig. 6.1 (b). The straight line is the graph of $\widetilde{w_k}$, the curved line is a possible graph of $\widetilde{f}$. The arrows show how to construct a homotopy F joining $\widetilde{f}$ to $\widetilde{w_k}$. The map $\exp \circ F$ is a homotopy joining f to w_k. $\square$

The lemma immediately implies items (0) and (2) above. Item (1) (injectivity) now follows from (0) and from the fact that the assignment $f \mapsto \widetilde{f}(2\pi)/2\pi$ is continuous, i.e., small changes in f result in small changes in the degree of f, but since the degree is an integer, sufficiently small changes in f result in no change at all in the degree of f! Finally, (3) (surjectivity) is obvious: given $k \in \mathbb{Z}$, for the appropriate f we take w_k.

The theorem is proved. $\square$

The geometric meaning of the degree of a map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ is "the number of times that the preimage circle wraps around the image circle". Thus the constant map $\mathbb{S}^1 \to 1 \in \mathbb{S}^1$ has degree 0 (the preimage circle wraps around the image circle zero times), the identity map has degree 1 (the preimage circle wraps around the image circle exactly once), the map w_{-17} has degree -17 (the preimage circle wraps around the image circle seventeen times in the negative direction (clockwise)).

Corollary 6.1. *The identity map of the circle is not homotopic to the constant map* $\mathbb{S}^1 \to 1 \in \mathbb{S}^1$.

Remark. The notion of degree of a map can be generalized from maps of the circle to itself to maps of the sphere $\mathbb{S}^n$ to itself for any n, and even to arbitrary n-dimensional oriented manifolds. Although the definition is not difficult, it is hard to prove in the general case (i.e., for any n) that the degree is well defined and depends only on the homotopy type of the map. To do that properly, you need *homology theory*, which lies outside the scope of this course.

6.4. A fixed point theorem

The theorem proved in the previous section has numerous important corollaries, several of which will eventually be discussed in this course. Here we only give one illustration, namely the Brouwer Fixed Point Theorem (for $n=2$). Other more general fixed point theorems lie at the basis of fundamental existence theorems in differential equations and their applications to engineering and especially economics (the so-called *Nash equilibrium*), but they require homology theory for their proofs.

Theorem 6.2 (Brouwer Fixed Point Theorem)**.** *Any continuous map of the* (*closed*) *disk has a fixed point, i.e., if* $f\colon \mathbb{D}^2 \to \mathbb{D}^2$ *is continuous, then there exists a point* $x \in \mathbb{D}^2$ *such that* $f(x) = x$.

For the proof, we will need a definition and a lemma. If A is a subspace of a topological space X, a continuous map $r\colon X \to A$ is said to be a *retraction* if r restricted to A is the identity. If a retraction $r\colon X \to A$ exists, then the subspace A is called a *retract* of X.

Lemma. *There is no retraction of the* 2*-disk on its boundary circle.*

Proof of the lemma. Suppose that there exists a retraction $r\colon \mathbb{D}^2 \to \partial\mathbb{D}^2$ of the 2-disk $\mathbb{D}^2$ on its boundary circle $\mathbb{S}^1 = \partial\mathbb{D}^2$. Consider the family $F_t(x)$ of maps $F_t\colon \mathbb{S}^1 \to \mathbb{S}^1$ given by the formula $F_t(e^{i\varphi}) = r(te^{i\varphi})$. The map F_0 is the constant map $\mathbb{S}^1 \to r(O)$ and the map F_1 (which is homotopic to F_0) is the identity map of the circle. This contradicts Corollary 6.1. ☐

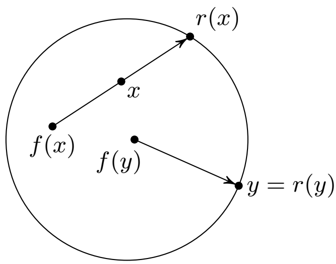


FIGURE 6.2. A retraction that does not exist

Proof of the theorem. To show that the Fixed Point Theorem follows from the lemma, assume that the theorem is false. For any $x \in D^2$, we have $f(x) \neq x$, and so the intersection point $r(x)$ of the ray $[f(x), x)$ with the boundary circle is well defined (look at Figure 6.2) and obviously the map $x \mapsto r(x)$ is a (continuous) retraction of $\mathbb{D}^2$ onto its boundary circle. But this contradicts the lemma. The theorem is proved. ☐

6.5. Exercises

6.1. If the restrictions of a map $f\colon X \to Y$ to its closed subsets $X_1, \dots X_k$, where $X_1 \cup \dots \cup X_k$ are all continuous, then f is continuous.

6.2. (a) Prove that if a map $f\colon X \to \mathbb{S}^1$ is not surjective, then f is homotopic to the constant map.

(b) Prove that if a map $f\colon X \to \mathbb{S}^n$ is not surjective, then f is homotopic to the constant map.

6.3. Prove that the 2-sphere with two points identified and the union of the 2-sphere with one of its diameters are homotopy equivalent.

6.4. Prove that the spaces $\mathbb{S}^1$ and $\mathbb{S}^1 \sqcup [0,1]/_\sim$, where $\sim$ denotes the identification of some point of $\mathbb{S}^1$ with the point $0 \in [0,1]$, are homotopy equivalent.

Here and below $X \vee Y$ denotes the *wedge* of two path connected spaces X and Y, i.e., the topological space obtained by identifying a point of X with a point of Y in the case when this topological space is well defined.

6.5. Prove that the sphere with g handles from which a point has been removed is homotopy equivalent to the space consisting of n circles passing through one point, and find n.

6.6. Prove that the spaces $\mathbb{S}^1 \vee \mathbb{S}^2$ and $\mathbb{R}^3 \setminus \mathbb{S}^1$ are homotopy equivalent.

6.7. Let X be the space $\mathbb{R}^3$ from which k copies of the circle have been removed (the circles are unknotted and unlinked, i.e., they lie in nonintersecting balls). Prove that X is homotopy equivalent to the wedge product of k copies of the space $\mathbb{S}^1 \vee \mathbb{S}^2$.

6.8. Let L be the union of two circles in $\mathbb{R}^3$ linked in the simplest way. Prove that $\mathbb{R}^3 \setminus L$ is homotopy equivalent to the wedge $\mathbb{S}^2 \vee \mathbb{T}^2$.

6.9. Prove that the following assertions are equivalent:

(1) any continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ has a fixed point;

(2) there is no retraction $r\colon \mathbb{D}^n \to \partial\mathbb{D}^n$;

(3) for any vector field v on $\mathbb{D}^n$ such that $v(x) = x$ for all $x \in \partial\mathbb{D}^n$, there exists a point $x \in \mathbb{D}^n$ such that $v(x) = 0$ (for $n = 2$ this assertion is called "теорема о макушке" in Russian and "hedgehog theorem" in English).

6.10. Prove that A is a retract of X if and only if any continuous map $f\colon A \to Y$ can be extended to X.

6.11. Prove that if any continuous map $f\colon X \to X$ has a fixed point and A is a retract of X, then any continuous map $g\colon A \to A$ has a fixed point.

6.12. Let $\mathbb{S}^\infty$ be the set of all points $(x_1, x_2, \ldots)$, $x_i \in \mathbb{R}$, such that only a finite number of x_i are nonzero and $\sum x_i^2 = 1$, supplied with the natural topology. Prove that the space $\mathbb{S}^\infty$ is *contractible* (i.e., homotopy equivalent to a point). *Hint*: Prove that the identity map is homotopic to the map $(x_1, x_2, \ldots) \mapsto (0, x_1, x_2, \ldots)$.

Lecture 7

Vector fields on the plane

The notion of vector field comes from mechanics and physics. Examples: the velocity field of the particles of a moving liquid in hydrodynamics, or the field of gravitational forces in Newtonian mechanics, or the field of electromagnetic induction in electrodynamics. In all these cases, a vector is given at each point of some domain in space, and this vector changes continuously as we move from point to point. In this lecture we will study a simpler model situation: vector fields on the plane (rather than in space).

In mathematics, the notion of *smooth* vector field is a basic notion of differential equations (analysis) and is not a topological notion. However, in this lecture we will consider the more general (topological!) notion of *continuous* vector field and show how the notion of degree of circle maps can be used in this context, and so can be very efficiently applied to differential equations.

7.1. Trajectories and singular points

A *vector field* V in the plane $\mathbb{R}^2$ is an assignment of a vector to each point of the plane. In the coordinates x, y of $\mathbb{R}^2$, it may be expressed as

$$X = \alpha(x, y), \quad Y = \beta(x, y),$$

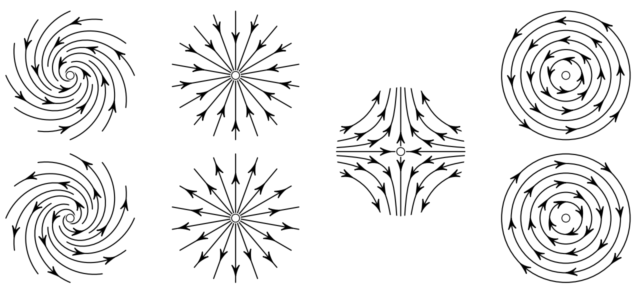
where $\alpha\colon \mathbb{R}^2 \to \mathbb{R}$ and $\beta\colon \mathbb{R}^2 \to \mathbb{R}$ are real-valued functions on the plane, (x, y) are the coordinates of the point p, and (X, Y) are the coordinates of the vector $V(p)$. If the functions α and β are continuous, then the vector field V is called *continuous*, and if α and β are smooth (infinitely differentiable), then V is called *smooth*. We will consider only continuous vector fields in what follows, and therefore omit the adjective "continuous".

A *singular point* p of a vector field V is a point where V vanishes: $V(p)=0$; when V is a velocity field, such a point is often called a *rest point*, when V is a field of forces, it is called an *equilibrium point*.

A *trajectory* of the vector field V through the point $p \in \mathbb{R}^2$ is a curve $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ passing through p and tangent at all its points to the vector field (more precisely, the vector $V(q)$ is equal to the derivative $d\gamma(t)/dt$ at each point $q \in C$). When we picture a vector field, instead of drawing numerous vectors in the plane, it is much clearer to draw its trajectories. It is a classical theorem in differential equations that *the trajectories of a smooth vector field always exist*. We do not need this theorem in the following theory, we need it only to motivate our illustrations, so will not prove it.

7.2. Generic singular points of plane vector fields

We will now define certain types of singular points of plane vector fields. To define these points, we will not write explicit formulas for the vectors of the field, but instead describe the picture of its trajectories near the singular point and give physical examples of such singularities (see Figure 7.1).



focus node saddle center

FIGURE 7.1. Singular points of vector fields

A *node* is a singular point contained in all the nearby trajectories; if all the trajectories move towards the point, the node is called *stable* and *unstable* if all the trajectories move away from the point. As an example, we can consider the gravitational force field of water droplets flowing down the surface $z=x^2+y^2$ near the point $(0,0,0)$ (stable node) or down the surface $z=-x^2-y^2$ near the same point (unstable node).

A *saddle* is a singular point containing two transversal trajectories, called *separatices*, one of which is ingoing, the other outgoing, the other trajectories behaving like a family of hyperbolas whose asymptotes are the separatrices. As an example, we can consider the gravitational force field of water droplets flowing down the surface $z = x^2 - y^2$ near the point $(0, 0, 0)$; here the separatrices are the bissectors of the coordinates axes Oxy.

A *center* is a singular point near which the trajectories behave like the family of concentric circles centered at that point; a center is called *positive* if the trajectories rotate counterclockwise and *negative* if they rotate clockwise. As an example, we can consider the velocity field obtained by rotating the plane about the origin with constant angular velocity.

A *focus* is a singular point that resembles the node, except that the trajectories, instead of behaving like the set of straight lines passing through the point, behave as a family of logarithmic spirals converging to it (*stable focus*) or diverging from it (*unstable focus*).

A singular point is called *generic* if it is of one of the three following types described above: node, saddle, focus. Note that the center is *not* generic. A vector field is called *generic* if it has a finite number of singular points all of which are generic.

Remark 7.1. Let us explain informally why the term generic is used here. Generic fields are, in fact, the "most general" ones in the sense that, first, they occur "most often" (i.e., as close as we like to any vector field there exists a generic one) and, second, they are "stable" (any vector field close enough to a generic one is also generic and has the same number of singular points). These statements are not needed in this course, so we will not make them more precise nor prove them.

Remark 7.2. It can be proved that the saddle and the center are *not* topologically equivalent to each other and not equivalent to the node or to the focus; however, the focus and the node *are* topologically equivalent; as topologists, we should not distinguish them, but we do, following the traditions of the theory of dynamical systems (where an equivalence relation stricter than homeomorphism is used). We do not use (an hence do not define) this relation.

7.3. The index of plane vector fields

Suppose a (continuous but not necessarily generic) vector field V in the plane is given. Let $\gamma(\mathbb{S}^1)$ be a closed curve in the plane (i.e., γ is an embedding ("вложение" in Russian) of $\mathbb{S}^1$ into $\mathbb{R}^2$) not passing through any singular points of V; let us denote $C := \gamma(\mathbb{S}^1)$. To each vector $V(c)$, $c \in C$, let us assign the unit vector of the same direction as $V(c)$ issuing from the origin of coordinates $O \in \mathbb{R}^2$; we then obtain a map $g\colon C \to \mathbb{S}^1_1$ (where $\mathbb{S}^1_1 \subset \mathbb{R}^2$ denotes the unit circle centered at O), called the *Gauss map* corresponding to the vector field V and to the curve γ. Now we define the *index of the vector field V along the curve γ* as the degree of the circle map $(g \circ \gamma)\colon \mathbb{S}^1 \to \mathbb{S}^1$:

$$\boxed{\operatorname{ind}(\gamma, V) := \deg(g \circ \gamma).}$$

Intuitively, the index is the total number of revolutions in the positive (counterclockwise) direction that the vector field performs when we go around the curve once.

Remark 7.3. A simple way of computing $\operatorname{ind}(\gamma)$ is to fix a ray not containing singular points issuing from O and count the number of times p the endpoint of $V(c)$ passes through the ray in the positive direction and the number of times q in the negative one; then $\operatorname{ind}(\gamma) = p - q$.

Theorem 7.1. *Suppose that a simple closed curve $C = \gamma(\mathbb{S}^1)$, $\gamma\colon \mathbb{S}^1 \hookrightarrow \mathbb{R}^2$, does not pass through any singular points of a vector field V and bounds a domain that also does not contain any singular points of V. Then*

$$\operatorname{ind}(\gamma, V) = 0.$$

To prove this theorem, we will need a stronger version of the Jordan Curve Theorem, known as the Schoenflies theorem, which we state as a fact without proof.

Fact (Schoenflies Theorem)**.** *Let $C := \gamma(S^1)$ be a closed curve in the plane. Then there exists a homeomorphism h of $\mathbb{R}^2$ that takes the domain D bounded by C to the unit disk centered at the origin O.*

Proof of Theorem 7.1. Let $h\colon D \to \mathbb{D}^2$ be a homeomorphism given by the Schoenflies theorem of the domain D to the unit disk centered at the origin O. Consider the family of all circles $\mathbb{S}^1_r$ of radius $r \leqslant 1$ centered at O. Obviously,

$$\operatorname{ind}(\gamma, V) = \operatorname{ind}(h^{-1}(\mathbb{S}^1_1), V).) \tag{*}$$

The vector $V(h^{-1}(O))$ is nonzero, hence for a small enough r_0, all the vectors $V(s)$, $s \in h^{-1}(\mathbb{S}^1_{r_0})$, differ little in direction from $V(h^{-1}(O))$, so that we have $\operatorname{ind}(h^{-1}(\mathbb{S}^1_{r_0}), V) = 0$. But then, by continuity,

$$\operatorname{ind}(h^{-1}(S^1_r), V) = 0 \quad \text{for all } r \leqslant 1.$$

Now the theorem follows from $(*)$. $\square$

Now suppose that V is a smooth plane vector field and p is a singular point of V. Let C be a circle centered at p such that no other singular points are contained in the disk bounded by C. Then the *index of V at the singular point p* is defined as $\operatorname{ind}(p, V) := \operatorname{ind}(C, V)$. This index is well defined, i.e., it does not depend on the radius of the circle C (provided that the disk bounded by C does not contain any other singular points); this follows from the next theorem.

Theorem 7.2. *Suppose that a simple closed curve γ does not pass through any singular points of a vector field V and bounds a domain that contains exactly one singular point a_0 of V. Then* $\operatorname{ind}(\gamma, V) = \operatorname{ind}(a_0, V)$.

The proof is similar to that of Theorem 7.1 and is left as an exercise.

7.4. Exercises

7.1. On the complex plane, consider the vector field $v(z) = z^n/|z|^{n-1}$ for $z \neq 0$, $v(0) = 0$. Find the index of the singular point of this field (for any integer n).

7.2. Prove that the index of the curve γ is equal to the sum of indices of the singular points that it encircles.

7.3. Suppose that two vector fields v and w are given on a closed non-self-intersecting curve in such a way that at any point X the vectors $v(X)$ and $w(X)$ do not point in exactly opposite directions. Prove that the indices of γ with respect to these vector fields are equal.

7.4*. Prove that any polynomial $P(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ with complex coefficients has at least one complex root.

7.5. Let us say that a vector field v is *even* if $v(x) = v(-x)$ and *odd* if $v(x) = -v(-x)$. Prove that the index of the point O for an even field is even and is odd for an odd field.

7.6. A closed self-intersecting curve divides the plane into several regions. By choosing a point O in each region, we can assign to the region the number of revolutions performed by the vector $\overrightarrow{OX}$ when the point X goes around the curve. Prove that if two regions have a common boundary, then the two numbers for the two regions differ by 1.

7.7*. On the boundary circles of an annulus consider a vector field without singular points such that the vectors are tangent to the circles and the vectors at any two corresponding points of the circles have opposite directions. Extend this vector field to a vector field without singular points on the entire annulus.

7.8. Let f be a smooth function on the plane. Prove that the index of an isolated singular point of the vector field $v = \operatorname{grad} f$

(a) can be equal to $1, 0, -1, -2, \ldots$ and

(b)* cannot be equal to the other integers.

Lecture 8

Vector fields on surfaces

In this lecture, we discuss vector fields on orientable surfaces. Here we will see that there is a deep relationship between the *global* topological properties of the surface and the structure of vector fields on it, namely the (*local*!) characteristics of its singular points. The previous lecture will serve as the *local* version of the theory.

8.1. What is a vector field on a surface?

A simple example of a vector field on a surface is the velocity field of points on the 2-sphere rotating with constant speed around the N-S axis. In order to define this notion mathematically in the general case, we will assume that our (compact closed orientable) surface M is embedded in $\mathbb{R}^3$. This means that M can be covered by a finite number of open disks $\{U_k\}$ each of which is the graph ("график" in Russian) of a univalent function $z_k = F_k(x_k, y_k)$ with respect to an orthonormal system of coordinates (O_k, x_k, y_k, z_k) (called *local coordinates*).

Thus *locally* the situation here is the same as in the previous lecture: one can define smooth vector fields, trajectories, singular points of a vector field, generic vector fields, the index of a vector field at a singular point, etc. However, *for an arbitrary curve* $\gamma\colon \mathbb{S}^1 \to M$*, the index of a vector field* $\operatorname{Ind}(V,\gamma)$ *cannot be correctly defined*, because the Gauss map uses the parallel shift of vectors to a common origin, and such a shift is not well defined on the whole surface. Nevertheless, Theorems 7.1 and 7.2 of the previous lecture remain valid *provided* that they are understood locally, i.e., as taking place in a disk $U_k \subset M$.

Remark. A more appropriate setting for this lecture is the framework of smooth surfaces (2-dimensional differentiable manifolds), where the vector field consists of vectors lying in the so-called "tangent planes" to the surface. Since this notion is not known to the listeners of this course, we have to resort the elementary approach given above, which involves no tangent planes.

The *index of a generic vector field* V on a closed compact orientable surface M is the sum of all indices for all singular points of this field (we denote it by $\mathrm{Ind}(M, V)$).

As for the case of a plane, a *generic vector field on a surface* M is defined as a generic vector field on all the U_k with a finite number of singular points, all of which are generic (i.e., are nodes, or foci, or saddles).

8.2. Two lemmas

The two following lemmas will be needed in the proof of the main result of this lecture, the Poincaré Theorem.

Lemma 8.1. *If p is a nonsingular point of a generic vector field V, D is a disk centered at p, and V_0 is any nonzero vector, then there exists another vector field W with the same singular points, coinciding with V outside of D and such that $W(p) = V_0$.*

Proof. By continuity, there is a disk D_0 concentric to D such that all the vectors $V(q)$, $q \in D_0$, have a direction that differs by less than 1° from the direction of $V(p)$. Let r be the radius of D_0, α be the angle between $V(p)$ and V_0, and S^1_s be the circle of radius $s \leqslant r$ centered at p. Then the required vector field W is obtained from V by rotating all the vectors $V(m)$, $m \in S^1_s$, by the angle $\alpha(r-s)/r$ and replacing $V(p)$ by V_0. $\square$

Lemma 8.2. *For any generic vector field V on a surface M, there is a triangulation of M such that any open 2-simplex contains no more than one singular point.*

Proof. Since the number of singular points is finite, by slightly moving the vertices of the triangulation, we can ensure that no singular point is a vertex or a point of an edge of the triangulation. By performing iterated barycentric subdivisions a sufficient number of times, we can ensure that there is no more than one singular point in each closed 2-simplex. Then we again slightly move the vertices of the triangulation so that no singular point is a vertex or lies on an edge. Then each singular point will lie inside a 2-simplex containing no other singular points. $\square$

8.3. The Poincaré index theorem

Henri Poincaré proved the following beautiful theorem, establishing a deep connection between the character of singular points of vector fields and the topology (as expressed by the Euler characteristic) of the surface on which they are defined.

Theorem 8.1. *The index of any smooth generic vector field on a (closed compact connected triangulated) orientable surface is equal to the Euler characteristic of this surface.*

Proof. The proof will be in two parts. In the first part, we will construct a special vector field whose index is indeed equal to the Euler characteristic of the surface. In the second part, we will prove that all generic vector fields on a given surface have the same index.

Part 1. Let us fix a triangulation of our surface M. We will construct a special continuous vector field on the triangulated surface with singular points at all the vertices, at the midpoints of all the edges, and at the barycenters of all the faces, such that the index of this vector field is equal to the Euler characteristic of M. At the midpoint of each edge, we place a saddle point whose ingoing separatrix goes along the edge and whose outgoing separatrix goes to the barycenters of the two triangular faces adjacent to the edge. At each vertex, we place an unstable node so that the edges issuing from the vertex are covered by outgoing trajectories of the node. At the barycenter of each face, we place a stable node so that its ingoing trajectories include the three separatrices coming to the barycenter from the three saddle points at the midpoints of the face's three sides (Fig. 8.1). Finally, it is easy to see that the vector fields already constructed in the neighborhoods of the three types of points (vertices, midpoints, barycenters) can be extended continuously so as to cover the entire surface.

The index of the vector field thus constructed is obviously equal to the Euler characteristic $\chi = V - E + F$ of the surface. Indeed, the nodes at the vertices and the baricenters have index equal to $+1$, so that the nodes contribute $V + F$ to the index, while the saddle points have index equal to -1, so they contribute $-E$, and all that adds up to χ.

Part 2. Let V_1 and V_2 be two generic vector fields on our surface; our aim is to prove that they have the same index. First, by using Lemma 8.2, we can assume that all the singular points of V_1 and V_2 lie inside the 2-simplices (triangles) of the triangulation, no more than one in each.

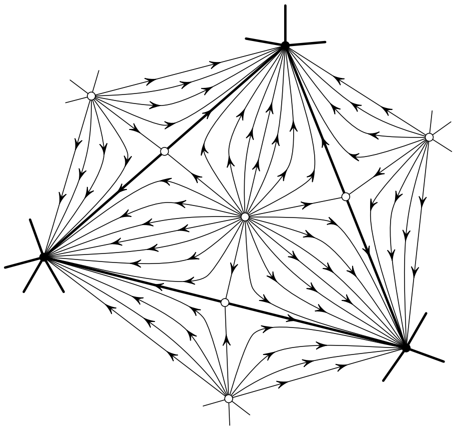


FIGURE 8.1. Singular points of the special vector field

Next, by applying Lemma 8.1 at each vertex, we can assume that the vectors $V_1(a)$ and $V_2(a)$ have the same direction at each vertex a.

Now let us fix an orientation of M. Then each edge ab acquires two opposite orientations, ab and ba, from the two faces adjacent to it. Let a mobile point x move from a to b and then back to a; as x moves from a to b, consider the rotation of the vector issuing from a and equal to $V_1(x)$ followed by the rotation of the vector issuing from a and equal to $V_2(x)$ as the point x moves back from b to a; denote by d_{ab} the number of revolutions performed by the vector (d_{ab} is a well defined integer, because the two vector fields coincide at the vertices). In a similar way, we can define d_{ba}. Obviously, $d_{ab} = -d_{ba}$. Summing over the set E of all edges, we obtain

$$\sum_{(ab)\in E} (d_{ab} + d_{ba}) = 0. \tag{$*$}$$

Next let us look at this sum from the point of view of the set F of faces. Let $(abc) \in F$, where the cyclic order a, b, c agrees with the chosen orientation of M. Now consider the sum $d_{ab} + d_{bc} + d_{ca}$; it does not change if we first perform the rotation of all the vectors V_1 and then of all the vectors V_2; therefore,

$$\begin{aligned} d_{ab} + d_{bc} + d_{ca} &= \mathrm{Ind}(\langle abc\rangle, V_1) + \mathrm{Ind}(\langle bac\rangle, V_2) \\ &= \mathrm{Ind}(\langle abc\rangle, V_1) - \mathrm{Ind}(\langle abc\rangle, V_2), \end{aligned} \tag{$**$}$$

where $\langle abc\rangle$ denotes the (positively oriented) closed curve bounding the face (abc). Rewriting the sum $(*)$ as a sum over the faces, using $(**)$, and Theorems 7.1 and 7.2, we obtain:

$$\begin{aligned} 0 &= \sum_{(abc)\in F} \big(\mathrm{Ind}(\langle abc\rangle, V_1) - \mathrm{Ind}(\langle abc\rangle, V_2)\big) \\ &= \sum_{(abc)\in F} \mathrm{Ind}(\langle abc\rangle, V_1) - \sum_{(abc)\in F} \mathrm{Ind}(\langle abc\rangle, V_2) \\ &= \mathrm{Ind}(M, V_1) - \mathrm{Ind}(M, V_2). \end{aligned}$$

The theorem is proved. □

8.4. Applications

Here we state only two immediate applications of Poincaré's Theorem (there will be more in the exercise classes).

Corollary 8.1. *Any generic smooth vector field on the sphere has at least two singular points.*

Corollary 8.2. *Any smooth force field on the configuration space of the pentagonal linkage with fixed hinges at the distance* 3.9 *from each other and* 4 *mobile sides of length* 1 *has at least two equilibrium points.*

8.5. Exercises

8.1. On the torus construct a vector field without singular points.

8.2. On the Klein bottle construct a vector field without singular points.

8.3. On the sphere construct a vector field with one nongeneric singular point.

8.4. On the projective plane construct a vector field with one singular point.

8.5. On the projective plane, does there exist a vector field (a) without any singular points, (b) with two singular points, both generic, (c) with three singular points, all generic, (d) with 17 singular points, all generic?

8.6. On the sphere with two handles construct a vector field with one singular point.

8.7. To each point X on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ a nonzero vector $v(X)$ in space is assigned. The vector depends continuously on the point of the sphere, but is not necessarily tangent to it. Prove that at least one of the vectors $v(X)$ is perpendicular to the tangent plane to the sphere at the point X.

8.8. Let $f\colon \mathbb{S}^2 \to \mathbb{S}^2$ be a continuous map. Prove that there exists a point $x \in \mathbb{S}^2$ such that $f(x) = \pm x$.

Lecture 9
Curves in the plane

In this lecture, we study curves and points lying in the plane $\mathbb{R}^2$ and introduce two important invariants: the Whitney index (or winding number) of a curve $w(\gamma)$ and the degree of a point with respect to a curve $\deg(p, \gamma)$. The Whitney index will allow us to classify curves immersed in the plane up to regular homotopy and the degree of a point with respect to a curve will help us prove the so-called "Fundamental Theorem of Algebra".

9.1. Regular curves and regular homotopy

A closed curve $f\colon \mathbb{S}^1 \to \mathbb{R}^2$ is called *regular* if it has a continuously changing nonzero tangent vector at each point; this means that for any $s \in \mathbb{S}^1$ there exists a neighborhood $U \subset \mathbb{S}^1$, $s \in U$, such that the restriction $f|_U$ defines the graph of a continuously differentiable function in some coordinate system in $\mathbb{R}^2$ and this graph has a nonzero tangent vector at the point $f(s)$. Note that a regular curve can have self-intersection points and even "overlaps", i.e., its image $f(\mathbb{S}^1)$ may contain intervals that are the image of disjoint intervals of $\mathbb{S}^1$, $f(U) = f(V)$, $U \cap V = \varnothing$.

A *regular homotopy* of a curve $f\colon \mathbb{S}^1 \to \mathbb{S}^2$ is a homotopy of that curve (i.e., a map $F\colon \mathbb{S}^1 \times [0,1] \to \mathbb{R}^2$ satisfying $F(s,0) = f(s)$ for all $s \in \mathbb{S}^1$) determining a regular curve for each $t \in [0,1]$ (i.e., the curve $F(s,t_0)$ is regular for any fixed $t_0 \in [0,1]$). Note that the "disappearance of a little loop", which can occur in a homotopy (see Fig. 9.1), is impossible in a *regular* homotopy (why?).

9.2. Immersed curves and regular homotopy

An immersed curve is a regular curve which is generic in the sense that its singular points cannot be destroyed by arbitrarily small changes. The exact definition is the following. A regular curve $f\colon \mathbb{S}^1 \to \mathbb{R}^2$ is said to be

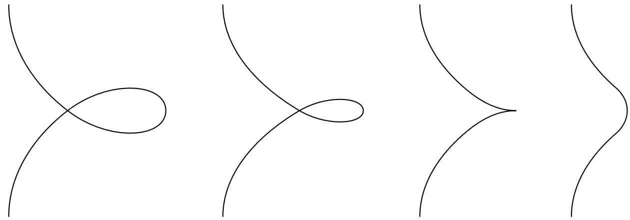


FIGURE 9.1. Disappearance of a little loop

an *immersion* if f is not a bijection at only a finite number of points d_j, and these points are *transversal double points*, i.e., their preimages are pairs of points and the two tangent vectors at each d_j are linearly independent.

Our aim is to classify immersed curves in the plane up to regular homotopy. This will be done by using an invariant defined in the next section.

9.3. The Whitney index

The *Whitney index* (also called *winding number*) $w(f)$ of a regular curve $f\colon \mathbb{S}^1 \to \mathbb{R}^2$ (not necessarily immersed) is defined as the degree of the Gauss map $\gamma_{df}\colon \mathbb{S}^1 \to \mathbb{S}^1$ determined by the tangent vector to the curve; this means that γ_{df} is obtained by parallel translation of the mobile tangent vector $df(\varphi)$ to the origin and normalizing it, and then letting φ vary from 0 to 2π.

There is a simple practical method for computing $w(f)$ for an immersed curve f: we consider all the horizontal tangent vectors to f and assume that there is a finite number of them, then we count the number of these vectors of different types and combine these numbers in the appropriate way. For the details, see the exercise class.

Clearly, *the Whitney index $w(f)$ is an invariant of regular homotopy* (because it is continuous and integer-valued).

9.4. Classification of immersed curves

In our classification we will ignore orientation, i.e., will not distinguish a curve f from the curve $f \circ \mathrm{sym}$, where sym is the symmetry of $\mathbb{S}^1$ with respect to a diameter. This classification is given by the following theorem.

Theorem 9.1 (H. Whitney, 1928)**.** *Any immersed curve (up to orientation) is regularly homotopic to exactly one of the following curves: the "figure eight curve", the circle, the circle with one small loop inside it, the circle with two small loops inside it, ..., the circle with* n *small loops inside it, and so on.*

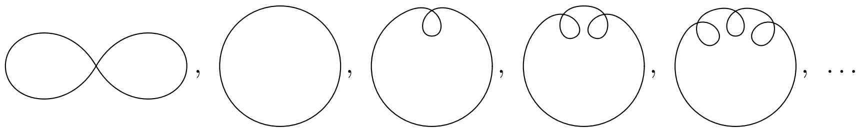


FIGURE 9.2. Classification of immersed curves

Proof. As usual for classification theorems, the proof is in two parts—one geometric, the other algebraic. In the geometric part, we construct a regular homotopy taking an arbitrary immersed curve to one of the curves listed in the theorem; we sketch this construction below (the details will be done in the exercise class). The second part consists in showing that the curves in the list are pairwise nonhomotopic; this is done by computing their Whitney indices; it turns out that they are all different (Exercise 9.10).

Let γ be the given immersed curve. We define a *simple loop* ω as a part of γ that starts and ends at a double point of γ and has no self-intersections (however, it can intersect other parts of γ).

First we prove that any immersed curve with self-intersections has a simple loop (Exercise 9.1).

Next we show that there is a homotopy after which all the simple loops do not intersect other parts of γ (Exercise 9.2).

Finally we use the homotopies shown in Figure 9.3 (Exercises 9.3 and 9.4) to conclude the proof of the theorem. □

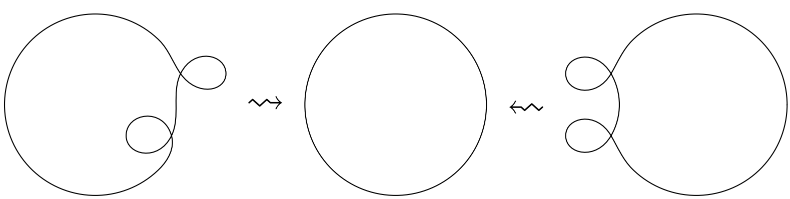


FIGURE 9.3. Two useful homotopies

Theorem 9.2 (The Whitney Theorem for the Sphere)**.** *Any immersed curve in the sphere is regularly homotopic to the circle or to the "figure eight curve".*

The proof is the subject of Exercise 9.6.

Note that here we do not classify "up to orientation" as in the previous theorem, but the classification "up to orientation" will be the same (why?). Concerning the proof, see the exercise class.

9.5. Degree of a point with respect to a curve

Consider a curve $f\colon \mathbb{S}^1 \to \mathbb{R}^2$ (not necessarily regular) and a point $p \in \mathbb{R}^2$ in its complement, $p \notin f(\mathbb{S}^1)$. Let φ be the angular parameter on $\mathbb{S}^1$ and V_φ be the vector joining the points p and $f(\varphi)$. As φ varies from 0 to 2π, the unit vector $V_\varphi/|V_\varphi|$ moves along the unit circle S_0 centered at p, defining a circle map $\gamma_f\colon S_0 \to S_0$. The *degree of the point p with respect to the curve f* is defined as the degree of the circle map γ, i.e., $\deg(p, f) := \deg(\gamma_f)$.

It is easy to prove that $\deg(p, f)$ does not change when p varies inside a connected component of $\mathbb{R}^2 \setminus f(S^1)$ (Exercise 9.7). If the point p is "far from" $f(S^1)$ (i.e., in the connected component of $\mathbb{R}^2 \setminus f(\mathbb{S}^1)$ with noncompact closure), then $\deg(p, f) = 0$ (Exercise 9.8).

Remark 9.1. There is a convenient method for computing the degree of any point p in the case when the curve f is immersed: join p by a (nonclosed) curve α not passing throuhg self-intersection points to a far away point a and move from a to p along that curve, adding one to the degree when you cross $f(\mathbb{S}^1)$ in the positive direction (i.e., so that the tangent vector looks to the right) and subtracting one when you cross it in the negative direction. The proof of the fact that you will always (independently of the choice of α) obtain $\deg(p, f)$ when your reach p is the object of Exercise 9.9.

9.6. The "fundamental theorem of algebra"

The Fundamental Theorem of Algebra says that *any polynomial*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0, \quad a_n \neq 0, \quad n > 0,$$

has at least one (*possibly complex*) *root*; here the coefficients a_i may be real or complex. We will prove this theorem assuming that $a_n = 1$ and $a_0 \neq 0$; this does not restrict generality (why?).

Consider the curve $f_n\colon \mathbb{S}^1 \to \mathbb{R}^2$ given by the formula $e^{i\varphi} \mapsto R_0^n e^{in\varphi}$, where R_0 is a (large) positive number that will be fixed later. Further, consider the family of curves $f_{p,R}\colon \mathbb{S}^1 \to \mathbb{R}^2$ given by the formula

$$e^{i\varphi} \mapsto p(Re^{i\varphi}), \quad \text{where } R \leqslant R_0.$$

We can assume that the origin O does not belong to $f_{p,R_0}(S^1)$ (otherwise the theorem is proved).

Lemma 9.1. *If R_0 is large enough, then* $\deg(O, f_{p,R_0}) = \deg(O, f_n) = n$.

Before proving the lemma, let us show that it implies the theorem.

By the lemma, $\deg(O, f_{p,R_0}) = n$. Let us continuously decrease R from R_0 to 0. If for some value of R the curve $f_{p,R}(\mathbb{S}^1)$ passes through the origin, the theorem is proved. So we can assume that $\deg(O, f_{p,R})$ changes continuously as $R \to 0$; but since the degree is an integer, it remains constant and equal to n. However, if R is small enough, the curve $f_{p,R}(\mathbb{S}^1)$ lies in a small neighborhood of a_0; but for such an R we have $\deg(O, f_{p,R}) = 0$. This is a contradiction, because $n \geqslant 1$.

It remains to prove the lemma. The equality $\deg(O, f_n) = n$ is obvious. To prove the other equality, it suffices to show that for any φ the difference Δ between the vectors $V_p(\varphi)$ and $V_n(\varphi)$ that join the origin O with the points $f_p(R_0e^{i\varphi})$ and $f_n(R_0e^{i\varphi})$, respectively, is small in absolute value (as compared to $R_0^n = |V_p(\varphi)|$) if R_0 is large enough. Indeed, by the definition of degree, if the mobile vector is replaced by another mobile vector whose direction always differs from the direction of the first one by less than $\pi/2$, the degree will be the same for the two vectors.

Clearly, $|\Delta| = |a_{n-1}z^{n-1} + \ldots + a_1z + a_0|$. Let us estimate this number, putting $z = R_0e^{\varphi}$ (we assume that $R_0 > 1$) and

$$A = \max\{a_{n-1}, a_{n-2}, \ldots, a_0\}.$$

We have

$$|a_{n-1}z^{n-1} + \ldots + a_1z + a_0| \leqslant |A(R_0^{n-1} + R_0^{n-2} + \ldots + 1)| \leqslant A \cdot n \cdot R_0^{n-1}.$$

Now if we put $R_0 := K \cdot A$, where K is a large positive number, we will obtain

$$|\Delta| \leqslant nA(KA)^{n-1} = nK^{n-1}A^n.$$

Let us compare this quantity to R_0^n; the latter equals $R_0^n = K^nA^n$, so for K large enough the ratio $|\Delta|/R_0^n$ is as small as we wish. This proves the lemma and concludes the proof of the theorem. $\square$

9.7. Exercises

9.1. Prove that any immersed curve with self-intersections has at least one simple loop.

9.2. Prove that for any simple loop ω of an immersed curve γ there exists a regular homotopy which changes only ω and replaces ω by a new simple loop that does not intersect other parts of γ.

9.3. Prove that the immersed curve shown on the left in Fig. 9.3 is regularly homotopic to the circle.

9.4. Prove that the immersed curve shown on the right in Fig. 9.3 is regularly homotopic to the circle.

9.5. Using the results of Exercises 9.1–9.4, prove the Whitney Theorem.

9.6. Prove the Whitney Theorem for the sphere.

9.7. Prove that $\deg(p, f)$ does not change when p varies inside a connected component of $\mathbb{R}^2 \setminus f(S^1)$.

9.8. Prove that if the point p is "far from" $f(S^1)$ (i.e., in the connected component of $\mathbb{R}^2 \setminus f(\mathbb{S}^1)$ with noncompact closure), then $\deg(p, f) = 0$.

9.9. Prove that the algorithm described in Remark 9.1 finds an integer d (which is independent of the choice of the curve α) and this integer is the degree: $\deg(p, f) = d$.

9.10. Compute the Whitney index of the curves shown in Figure 9.2.

Lecture 10

The fundamental group

The fundamental group is one of the most important invariants of homotopy theory. It also has numerous applications outside of topology, especially in complex analysis, algebra, theoretical mechanics, and mathematical physics. In our course, it will be the first example of a functor, assigning a group to each pathconnected topological space and a group homomorphism to each continuous map of such spaces, thus reducing topological problems about spaces to problems about groups, which can often be effectively solved.

10.1. Main definitions

Let M be a topological space with a distinguished point $p \in M$. A curve $c\colon [0,1] \to M$ such that $c(0) = c(1) = p$ will be called a *loop* with *basepoint* p. Two loops c_0, c_1 with basepoint p are called *homotopic rel endpoints* if there is a homotopy $F\colon X \times [0,1] \to Y$ joining c_0 to c_1 such that $F(t,x) = p$ for all $t \in [0,1]$.

Two curves c_0, c_1 such that $c_0(x) = c_1(x) = p$ (not necessarily loops) are called *homotopic rel* p if there is a homotopy H joining c_0 to c_1 such that $H(t,x) = p$ for all $t \in [0,1]$.

If c_1 and c_2 are two loops with basepoint p, then the loop $c_1 \cdot c_2$ given by

$$c_1 \cdot c_2(t) := \begin{cases} c_1(2t) & \text{if } t \leqslant \dfrac{1}{2}, \\ c_2(2t-1) & \text{if } t \geqslant \dfrac{1}{2}. \end{cases}$$

is called the *product* of c_0 and c_1.

Proposition 10.1. *Classes of loops homotopic rel endpoints form a group with respect to the product operation induced by* $\cdot$.

Proof. First notice that the operation is indeed well defined on homotopy classes. For, if the paths c_1, c_2 are homotopic to $\widetilde{c}_1$, $\widetilde{c}_2$, respectively, via the maps $h_i\colon [0,1]\times[0,1]\to M$, $i=1,2$, then the map h, defined by

$$h(t,s) := \begin{cases} h_1(2t,s) & \text{if } t \leqslant \frac{1}{2}, \\ h_2(2t-1,s) & \text{if } t \geqslant \frac{1}{2} \end{cases}$$

is a homotopy rel endpoints joining $c_1\cdot c_2$ to $\widetilde{c_1}\cdot\widetilde{c_2}$.

Obviously, the role of the unit is played by the homotopy class of the constant map $c_0(t)=p$. Then the inverse to c will be the homotopy class of the map $c'(t) := c(1-t)$. What remains is to check the associative law: $(c_1\cdot c_2)\cdot c_3$ is homotopic rel p to $c_1\cdot(c_2)\cdot c_3)$ and to show that $c\cdot c'$ is homotopic to c_0. In both cases the homotpy is done by a reparametrization in the preimage, i.e., on the square $[0,1]\times[0,1]$.

For associativity, consider the following continuous map ("reparametrization") of the square into itself

$$R(t,s) = \begin{cases} (t(1+s),s) & \text{if } 0 \leqslant t \leqslant 1/4, \\ (t+s/4,s) & \text{if } 1/4 \leqslant t \leqslant 1/2, \\ (1-1/(1+s)+t/(1+s),s) & \text{if } 1/2 \leqslant t \leqslant 1. \end{cases}$$

Then the map $c_1\cdot(c_2\cdot c_3)\circ R\colon [0,1]\times[0,1]\to M$ provides a homotopy rel endpoints joining the loops $c_1\cdot(c_2\cdot c_3)$ and $(c_1\cdot c_2)\cdot c_3$.

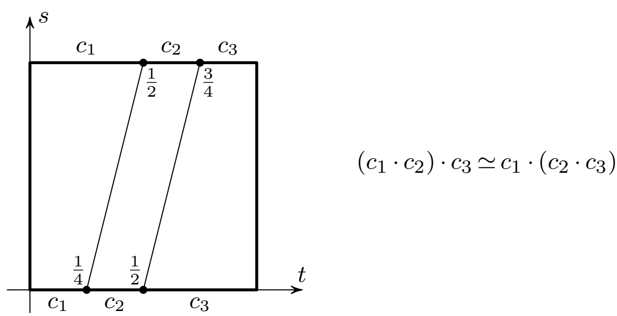


FIGURE 10.1. Associativity of multiplication

Similarly, a homotopy joining $c\cdot c'$ to c_0 is given by $c\cdot c'\circ I$, where the reparametrization $I\colon [0,1]\times[0,1]\to[0,1]\times[0,1]$ is defined as

$$I(t,s) = \begin{cases} (t,s) & \text{if } 0 \leqslant t \leqslant (1-s)/2, \text{ or } (1+s)/2 \leqslant t \leqslant 1, \\ ((1-s)/2,s) & \text{if } (1-s)/2 \leqslant t \leqslant (1+s)/2, \end{cases}$$

Notice that while the reparametrization I is discontinuous along the wedge $t=(1\pm s)/2$, the map $(c\cdot c')\circ I$ is continuous by the definition of c'. $\square$

The group described in Proposition 10.1 is called the *fundamental group* of M at p and is denoted by $\pi_1(M, p)$.

It is natural to ask to what extent $\pi_1(M, p)$ depends on the choice of the point $p \in M$. The answer is given by the following proposition.

Proposition 10.2. *If p and q belong to the same path connected component of M, then the groups $\pi_1(M, p)$ and $\pi_1(M, q)$ are isomorphic.*

Proof. Let $\rho\colon [0, 1] \to M$ be a path connecting the points p and q. It is natural to denote the path $\rho \circ S$, where $S(t) = 1 - t$, by ρ^{-1}. It is also natural to extend the "$\cdot$" operation to paths with different initial points and endpoints provided the endpoint of the first factor coincides with the initial point of the second one. With these conventions established, let us associate to a path $c\colon [0, 1] \to M$ with $c(0) = c(1) = p$ the path $c' := \rho^{-1} \cdot c \cdot \rho$ with $c'(0) = c'(1) = q$. In order to finish the proof, we must show that this correspondence takes paths homotopic rel p to paths homotopic rel q, respects the group operation and is bijective up to homotopy. These statements are proved by using appropriate rather natural reparametrizations, as in the proof of Proposition 10.1. $\square$

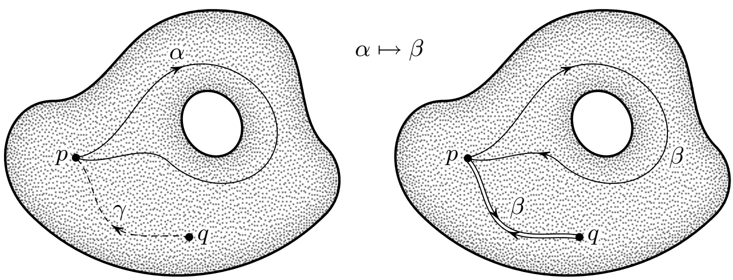


FIGURE 10.2. Change of basepoint isomorphism

Remark 10.1. Note that the isomorphism in Proposition 10.2 is not canonical: it follows from the construction that different choices of the connecting path ρ will produce isomorphisms between $\pi_1(M, p)$ and $\pi_1(M, q)$ which differ by an inner automorphism of either group.

If the space M is path connected, then the fundamental groups at all of its points are isomorphic and one simply talks about the *fundamental group of M* and often omits the basepoint from its notation: $\pi_1(M)$.

The *free* homotopy classes of curves (i.e., with no fixed base point) correspond exactly to the conjugacy classes of curves modulo changing base point, so there is a natural bijection between the classes of freely homotopic closed curves and conjugacy classes in the fundamental group.

A path connected space with trivial fundamental group is said to be *simply connected* (or sometimes 1-*connected*).

Remark 10.2. Since the fundamental group is defined modulo homotopy, it is the same group for homotopically equivalent spaces, i.e., *the fundamental group* $\pi_1(M)$ *is a homotopy invariant.*

10.2. Functoriality

Now suppose that X and Y are path connected, $f\colon X \to Y$ is a continuous maps with and $f(p) = q$. Let $[c]$ be an element of $\pi_1(X, p)$, i.e., the homotopy class rel endpoints of some loop $c\colon [0, 1] \to X$. Denote by $f_{\#}(c)$ the loop in (Y, q) defined by $f_{\#}(t) := f(c(t))$ for all $t \in [0, 1]$.

Proposition 10.3. *The assignment* $c \mapsto f_{\#}(c)$ *is well defined on classes of loops and determines a homomorphism* (*still denoted by* $f_{\#}$) *of fundamental groups:*

$$f_{\#}\colon \pi_1(X, p) \to \pi_1(Y, q)$$

(*called the* homomorphism induced by f), *which possesses the following properties* (*called* functorial):

- $(f \circ g)_{\#} = f_{\#} \circ g_{\#}$ (*covariance*);
- $(\mathrm{id}_X)_{\#} = \mathrm{id}_{\pi_1(X,p)}$ (*identity maps induce identity homomorphisms*).

The fact that the construction of an invariant (here the fundamental group) is functorial is very convenient for applications, as seen in the following example.

Example 10.1. Let us give another proof of the Brouwer fixed point theorem for the disk by using the isomorphisms $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ and $\pi_1(\mathbb{D}^2) = 0$ (which will be established later) and the functoriality of $\pi_1(\cdot)$.

We will prove (by contradiction) that there is no retraction of $\mathbb{D}^2$ on its boundary $\mathbb{S}^1 = \partial\mathbb{D}^2$. Let $r\colon \mathbb{D}^2 \to \mathbb{S}^1$ be such a retraction, let $i\colon \mathbb{S}^1 \to \mathbb{S}^2$ be the inclusion; choose a basepoint $x_0 \in \mathbb{S}^1 \subset \mathbb{S}^2$. Note that for this choice of basepoint we have $i(x_0) = r(x_0) = x_0$. Consider the sequence of induced maps:

$$\pi_1(\mathbb{S}^1, x_0) \xrightarrow{i_*} \pi_1(\mathbb{D}^2, x_0) \xrightarrow{r_*} \pi_1(\mathbb{S}^1, x_0).$$

In view of the isomorphisms noted above, this sequence is actually

$$\mathbb{Z} \longrightarrow 0 \longrightarrow \mathbb{Z}.$$

But such a sequence is impossible, because by functoriality we have

$$r_* \circ i_* = (r \circ i)_* = \mathrm{id}_* = \mathrm{id}_{\mathbb{Z}}\,.$$

The fundamental group behaves nicely with respect to Cartesian products, as the following proposition shows.

Proposition 10.4. *If X and Y are path connected spaces, then*

$$\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y).$$

Proof. Let us construct an isomorphism of $\pi_1(X) \times \pi_1(Y)$ onto $\pi_1(X \times Y)$. Let x_0, y_0 be the basepoints in X and Y, respectively. For the basepoint in $X \times Y$, let us take the point (x_0, y_0). Now to the pair of loops α and β in X and Y let us assign the loop $\alpha \times \beta$ given by $\alpha \times \beta(t) := (\alpha(t), \beta(t))$. The verification of the fact that this assignment determines a well-defined isomorphism of the appropriate fundamental groups is quite straightforward. For example, to prove surjectivity, for a given loop γ in $X \times Y$ with basepoint (x_0, y_0), we consider the two loops $\alpha(t) := (\mathrm{pr}_X \circ \gamma)(t)$ and $\beta(t) := (\mathrm{pr}_Y \circ \gamma)(t)$, where pr_X and pr_Y are the projections on the two factors of $X \times Y$. □

Corollary 10.1. *If C is contractible, then $\pi_1(X \times C) = \pi_1(X)$.*

The proof is an exercise.

10.3. The Seifert–van Kampen theorem

In this section, we state without proof a classical theorem which relates the fundamental group of the union of two spaces with the fundamental groups of the summands and of their intersection. The result turns out to give an efficient method for computing the fundamental group of a "complicated" space by putting it together from "simpler" pieces.

In order to state the theorem, we need a purely algebraic notion from group theory.

Let G_i, $i = 1, 2$, be groups, and let $\varphi_i \colon K \to G_i$, $i = 1, 2$ be monomorphisms. Then the *free product with amalgamation* of G_1 and G_2 with respect to φ_1 and φ_2, denoted by $G_1 *_K G_2$ is the quotient group of the free product $G_1 * G_2$ by the normal subgroup generated by all elements of the form $\varphi_1(k)(\varphi_2(k))^{-1}$, $k \in K$.

Theorem 10.1 (Van Kampen's Theorem)**.** *Let the path connected space X be the union of two path connected spaces A and B with path connected intersection containing the basepoint $x_0 \in X$. Let the inclusion homomor-*

phisms

$$\varphi_A\colon \pi_1(A\cap B)\to\pi_1(A),\quad \varphi_A\colon \pi_1(A\cap B)\to\pi_1(B)$$

be injective. Then $\pi_1(X, x_0)$ *is the amalgamated product*

$$\boxed{\pi_1(X, x_0)\cong\pi_1(A, x_0) *_{\pi_1(A\cap B, x_0)} \pi_1(B, x_0)}$$

10.4. Exercises

10.1. Prove that if C is contractible, then $\pi_1(C)=0$.

10.2. Prove that for any path connected topological space X we have $\pi_1(\mathrm{Cone}(X))=0$.

10.3. Prove that the fundamental group of the wedge product of n circles is isomorphic to the free group with n generators.

10.4. Prove that the group $\pi_1(n\mathbb{T}^2)$ is generated by elements $a_1, b_1, \ldots,$ a_n, b_n obeying to the unique relation

$$\prod_{i=1}^{n}(a_ib_ia_i^{-1}b_i^{-1}) = 1.$$

10.5. Prove that the group $\pi_1(n\mathbb{R}P^2)$ is generated by elements $a_1, \ldots,$ a_n, obeying to the unique relation $a_1^2\ldots a_n^2=1$.

10.6. (a) Prove that if $G=\pi_1(n\mathbb{T}^2)$, then $G/G'\cong\mathbb{Z}^{2n}$. (Here G' is the *commutant*, i.e. G' is the subgroup generated by all elements of the form $aba^{-1}b^{-1}$ for $a, b\in G$.)

(b) Prove that if $G=\pi_1(n\mathbb{R}P^2)$, then $G/G'\cong\mathbb{Z}^{n-1}\oplus\mathbb{Z}_2$.

10.7. Prove that $\pi_1(\mathbb{S}^n)=0$ for $n\geqslant 2$.

10.8. Prove that $\pi_1(\mathbb{C}P^n)=0$.

10.9. Prove that the fundamental group of the surface $n\mathbb{T}^2$ with $k\geqslant 1$ deleted discs is the free group of rank $2n+k-1$.

10.10. Prove that the fundamental group of the surface $n\mathbb{R}P^2$ with $k\geqslant 1$ deleted discs is the free group of rank $n+k-1$.

10.11. Suppose that X is the Möbius band, A is its boundary. Prove that A is not a retract of X.

10.12. Prove that any finite and connected CW-space is homotopy equivalent to a CW-space with only one vertex e^0.

Lecture 11
Covering spaces

A covering space (or covering) is not a space, but a mapping of spaces (usually manifolds) which, locally, is a homeomorphism, but globally may be quite complicated. The simplest nontrivial example is the exponential map $\mathbb{R} \to \mathbb{S}^1$ discussed in Lecture 6.

11.1. Definition and examples

In this lecture, we will consider only path connected spaces with basepoint and only basepoint-preserving continuous maps (not necessarily cellular). Suppose E, B are path connected topological spaces $p\colon E \to B$ is a continuous map such that $p^{-1}(y)$ is a discrete subspace, the cardinality of the set $p^{-1}(y) := D$ is independent of $y \in B$ and every $x \in p^{-1}(y)$ has a neighborhood on which p is a homeomorphism onto a neighborhood of $y \in B$, then the quadruple (p, T, B, D) is called a *covering* with *covering projection* p, *total space* E, *base* B, and *fiber* $D = p^{-1}(y)$.

If $n = |D|$ is finite, then (p, E, B, D) is said to be an *n-fold covering*. If D is countably infinite, we say that $p\colon E \to B$ is a *countable* covering.

Examples 11.1. (i) the map $w_3 : \mathbb{S}^1 \to \mathbb{S}^1$, given by $e^{i\varphi} \mapsto e^{i3\varphi}$ is a 3-fold covering of the circle by the circle;

(ii) the exponential map $\exp\colon \mathbb{R} \to \mathbb{S}^1$ is a countable covering of the circle by the real line;

(iii) the map $u\colon \mathbb{R}^2 \to \mathbb{T}^2$, $(x, y) \mapsto (2\pi\{x\}, 2\pi\{y\})$, where $\{\cdot\}$ denotes the fractional part of a real number, is a countable covering of the torus by the plane;

(iv) the map $\tau : \mathbb{S}^2 \to \mathbb{R}P^2$ obtained by identifying antipodal points of the sphere is a 2-fold covering of the projective plane.

Like any other important class of mathematical objects, covering spaces form a category. In this category, a morphism between two covering spaces $p_i\colon E_i \to B_i$, $i = 1, 2$, are pairs of (continuous, basepoint-preserving) maps $\varphi\colon B_1 \to B_2$ and $\Phi\colon E_1 \to E_2$ such that the following diagram is commutative:

$$\begin{tikzcd}
E_1 \arrow[r, "\Phi"] \arrow[d, "p_1"] & E_2 \arrow[d, "p_2"] \\
B_1 \arrow[r, "\varphi"] & B_2
\end{tikzcd}$$

Compositions of morphisms and identical morphisms are defined in the natural way. Then, obviously, an isomorphism of covering spaces is a morphism for which Φ and φ are homeomorphisms. Isomorphic covering spaces are considered identical.

If E is simply connected, then the covering $p\colon E \to B$ is called *universal*.

If $f\colon X \to B$ is continuous and $\widetilde{f}\colon X \to E$ satisfies $f = p \circ F$, then $\widetilde{f}$ is said to be a *lift* of f. The figure below shows the lift of a closed curve.

A homeomorphism of the total space of a covering E of E is called a *deck transformation* ("монодромия" in Russian), if it is a lift of the identity on B.

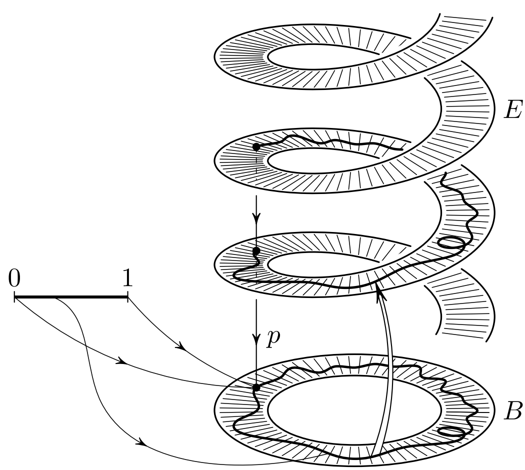


Figure 11.1. Lift of a closed curve

11.2. Path lifting and covering homotopy

In this section, we prove two important technical assertions which allow, given a covering space $p\colon E \to B$, to lift "upstairs" (i.e., to E) continuous processes taking place "downstairs" (i.e., in B). The underlying idea has already been exploited when we defined the degree of circle maps by using the exponential map (see Lecture 6), and we will now be generalizing that idea from the case of the exponential map to arbitrary covering spaces.

Lemma 11.1 (Path lifting lemma). *Any path in the base B of a covering space $p\colon E \to B$ can be lifted to the total space of the covering, and the lift is unique if its initial point in the covering is specified. More precisely, if $p\colon E \to B$ is a covering space, $\alpha\colon [0,1] \to B$ is any path, and $x_0 \in p^{-1}(\alpha(0))$, then there exists a unique map $\widetilde{\alpha}\colon [0,1] \to X$ such that $p \circ \widetilde{\alpha} = \alpha$ and $\widetilde{\alpha}(0) = x_0$.*

Proof. By the definition of covering space, for each point $b \in \alpha([0,1])$ there is a neighborhood U_b whose inverse image under p falls apart into disjoint neighborhoods each of which is projected homeomorphically by p onto U_b. The set of all such U_b covers $\alpha([0,1])$ and, since $\alpha([0,1])$ is compact, it possesses a finite subcover that we denote by $U_0, U_1, \dots U_k$.

Without loss of generality, we assume that U_0 contains $b_0 := \alpha(0)$ and denote by $\widetilde{U}_0$ the component of $p^{-1}(U_0)$ that contains the point x_0. Then we can lift a part of the path α contained in U_0 to $\widetilde{U}_0$ (uniquely!) by means of the inverse to the homeomorphism between $\widetilde{U}_0$ and U_0.

Now, again without loss of generality, we assume that U_1 intersects U_0 and contains points of $\alpha[0,1]$ not lying in U_0. Let $b_1 \in \alpha([0,1])$ be a point contained both in U_0 and U_1 and denote by $\widetilde{b}_1$ the image of b_1 under $p^{-1}\big|_{U_0}$. Let $\widetilde{U}_1$ be the component of the inverse image of U_1 containing $\widetilde{b}_1$. We now extend the lift of our path to its part contained in U_1 by using the inverse of the homeomorphism between $\widetilde{U}_1$ and U_1. Note that the lift obtained is the only possible one. Our construction in the case when the path is closed (i.e., is a loop) is shown in Figure 11.1.

Continuing in this way, after a finite number of steps we will have lifted the entire path $\alpha([0,1])$ to X, and the lift obtained will be the only one obeying the conditions of the lemma. □

Remark 11.1. Note that the lift of a closed path is *not* necessarily a closed path, as we have already seen in our discussion of the degree of circle maps.

Note also that if all paths (i.e., maps of $A=[0,1]$) can be lifted, it is not true that all maps of *any* space A can be lifted (see Exercise 11.20).

Now we generalize the path lifting lemma to homotopies, having in mind that a path is actually a homotopy, namely a homotopy of the one-point space. This trivial observation is not only the starting point of the formulation of the covering homotopy theorem, but also the key argument in its proof.

Theorem 11.1 (Covering homotopy theorem). *Any homotopy in the base of a covering space can be lifted to the covering, and the homotopy is unique if its initial map in the covering is specified as a lift of the initial map of the given homotopy. More precisely, if $p\colon E\to B$ is a covering, $F\colon A\times[0,1]\to B$ is any homotopy whose initial map $f_0(\cdot):=F(\cdot,0)$ possesses a lift $\widetilde{f_0}$, then there exists a unique homotopy $\widetilde{F}\colon A\times[0,1]\to X$ such that $p\circ\widetilde{F}=F$ and $\widetilde{F}(\cdot,0)=\widetilde{f_0}(\cdot)$.*

Proof. The theorem will be proved by reducing the theorem to the path lifting lemma from the previous section. Fix some point $\alpha\in A$. Define $\alpha_a(t):=F(a,t)$ and denote by x_a the point $\widetilde{f_0}(a)$. Then α_a is a path, and by the path lifting lemma, there exists a unique lift $\widetilde{\alpha}_a$ of this path such that $\widetilde{\alpha}(0)=x_a$. Now consider the homotopy defined by

$$\widetilde{F}(a,t) := \widetilde{\alpha}_a(t), \quad \text{for all } a\in A,\ t\in[0,1].$$

Then, we claim that $\widetilde{F}$ satisfies all the conditions of the theorem, i.e., $\widetilde{F}$ is continuous and unique. We leave this verification to the reader. □

Remark 11.2. The covering homotopy theorem is not true if $E\to B$ is an arbitrary surjection (and not a covering space). For a counterexample, see Exercise 11.7.

11.3. Role of the fundamental group

The projection p of a covering space $p\colon E\to B$ induces a homomorphism $p_\#\colon \pi_1(E)\to\pi_1(B)$. We will see that when the spaces E and B are "locally nice", the homomorphism $p_\#$ entirely determines (up to isomorphism) the covering space p over a given B. (What we mean by "locally nice" will be explained below.)

More precisely, in this section we will show that, provided that the "local nicety" condition holds, $p_\#$ is a monomorphism and that, given a subgroup G of $\pi_1(B)$, we can effectively construct a unique space E and a unique (up to isomorphism) covering map $p\colon E\to B$ for which G is the image of $\pi_1(E)$ under $p_\#$. Moreover, we will prove that there is a bijection between conjugacy classes of subgroups of $\pi_1(B)$ and isomorphism classes of coverings, thus achieving the classification of all coverings over a given base B in terms of $\pi_1(B)$.

Theorem 11.2. *The homomorphism $p_\#\colon \pi_1(E) \to \pi_1(B)$ induced by any (not necessarily locally nice) covering space $p\colon E \to B$ is a monomorphism.*

Proof. The theorem is an immediate consequence of the covering homotopy theorem proved in the previous section. Indeed, it suffices to prove that a nonzero element $[\alpha]$ of $\pi_1(E)$ cannot be taken to zero by $p_\#$. Assume that $p_\#([\alpha]) = 0$. This means that the loop $p \circ \alpha$, where $\alpha \in [\alpha]$, is homotopic to a point in B. By the homotopy lifting theorem, we can lift this homotopy to E, which means that $[\alpha] = 0$. $\square$

Now we describe the main construction of this lecture: given a space and a subgroup of its fundamental group, we will construct the corresponding covering. This construction works provided the space considered is "locally nice" in a sense that will be specified below.

Theorem 11.3. *For any "locally nice" space B and any subgroup $G \subset \subset \pi_1(B, b_0)$, there exists a unique covering space $p\colon X \to B$ such that $p_\#(X) = G$.*

Proof. The theorem is proved by means of another trick. Let us consider the set $P(B, b_0)$ of all paths in B issuing from b_0. Two paths $\alpha_i\colon [0, 1] \to B$, $i = 1, 2$ will be identified (notation $\alpha_1 \sim \alpha_2$) if they have a common endpoint and the loop λ given by

$$\lambda(t) = \begin{cases} \alpha_1(2t) & \text{if } 0 \leqslant t \leqslant 1/2, \\ \alpha_2(2-2t) & \text{if } 1/2 \leqslant t \leqslant 1. \end{cases}$$

determines an element of $\pi_1(B)$ that belongs to G. (The loop λ can be described as first going along α_1 (at double speed) and then along α_2 from its endpoint back to b_0, also at double speed.)

Denote by $X := P(B, b_0)/_\sim$ the quotient space of $P(B, b_0)$ by the equivalence relation just defined. Endow X with the "natural" topology (the formal definition is given below) and define the map $p\colon X \to B$ by stipulating that it takes each equivalence class of paths in $P(B, b_0)$ to the endpoint of one of them (there is no ambiguity in this definition, because equivalent paths have the same endpoint).

Then $p\colon X \to B$ is the required covering space. It remains to:
(o) define the topology on X; (i) prove that p is continuous; (ii) prove that p is a local homeomorphism; (iii) prove that $p_\#(\pi_1(X))$ coincides with G;

(iv) prove that p is unique. We will do this after defining what we mean by "locally nice".

Remark 11.3. To understand the main idea of the construction described above, the reader should try applying it in the case $G=0$ (construction of the universal cover).

Remark 11.4. The above construction is not effective at all, and cannot be used to describe the covering space obtained. However, in reasonably simple cases it is easy to guess what the space X is from the fact that the fundamental group of X is G and p is a local homeomorphism.

A topological space X is called *locally path connected* if for any point $x \in X$ and any neighborhood U of x there exists a smaller neighborhood $V \subset U$ of x which is path connected. A topological space X is called *locally simply connected* if for any point $x \in X$ and any neighborhood U of x there exists a smaller neighborhood $V \subset U$ of x which is simply connected.

Examples 11.2. (a) Let $X \subset \mathbb{R}^2$ be the union of the segments

$$\{(x, y)\colon y = 1/2^n, 0 \leqslant x \leqslant 1\} \quad n = 0, 1, 2, 3, \dots$$

and the two unit segments $[0, 1]$ of the x-axis and y-axis (see Figure 11.2 (a)). Then X is path connected but not locally path connected (at all points of the interval $(0, 1]$ of the x-axis).

(b) Let $X \subset \mathbb{R}^2$ be the union of the circles

$$\{(x, y)\colon x^2 + (y - 1/n)^2 = 1/n^2\} \quad n = 1, 2, 3, \dots;$$

the circles are all tangent to the x-axis and to each other at the point $(0, 0)$ (see Figure 11.2 (b)). Then X is path connected but not locally simply connected (at the point $(0, 0)$).

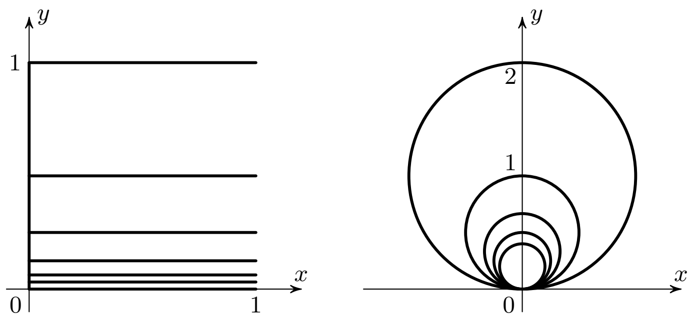


FIGURE 11.2. Not locally connected and not locally simply connected spaces

We will now conclude the proof of Theorem 11.3, assuming that B is locally path connected and locally simply connected.

(o) *Definition of the topology in* $X = P(B, b_0)/_\sim$. In order to define the topology, we will specify a base of open sets of rather special form, which will be very convenient for our further considerations. Let U be an open set in B and $x \in X$ be a point such that $p(x) \in U$. Let α be one of the paths in x with initial point x_0 and endpoint x_1. Denote by (U, x) the set of equivalence classes (with respect to $\sim$) of extensions of the path α whose segments beyond x_1 lie entirely inside U. Clearly, (U, x) does not depend on the choice of $\alpha \in x$.

We claim that (U, x) actually does not depend on the choice of the point x in the following sense: if $x_2 \in (U, x_1)$, then $(U, x_1) = (U, x_2)$. To prove this, consider the points $b_1 := p(x_1)$ and $b_2 := p(x_2)$. Join the points b_1 and b_2 by a path (denoted by β) contained in U.

Let $\alpha\alpha_1$ denote an extension of α, with the added path segment α_1 contained in U. Now consider the path $\alpha\beta\beta^{-1}\alpha_1$, which is obviously homotopic to $\alpha\alpha_1$. On the other hand, it may be regarded as the extension (beyond x_2) of the path $\alpha\beta$ by the path $\beta^{-1}\alpha_{-1}$. Therefore, the assignment $\alpha\alpha_1 \mapsto \alpha\beta\beta^{-1}\alpha_1$ determines a bijection between (U, x_1) and (U, x_2), which proves our claim.

Now we can define the topology in X by taking for a base of the topology the family of all sets of the form (U, x). To prove that this defines a topology, we must check that that a nonempty intersection of two elements of the base contains an element of the base. Let the point x belong to the intersection of the sets (U_1, x_1) and (U_2, x_2). Denote $V := U_1 \cap U_2$ and consider the set (V, x); this set is contained in the intersection of the sets (U_1, x_1) and (U_2, x_2) (in fact, coincides with it) and contains x, so that $\{(U, x)\}$ is indeed a base of a topology on X.

(i) *The map p is continuous.* Take $x \in X$. Let U be any path connected and simply connected neighborhood of $p(x)$ (it exists by the condition imposed on B). The inverse image of U under p is consists of basis open sets of the topology of X (see item (o)) and is therefore open, which establishes the continuity at an (arbitrary) point $x \in X$.

(ii) *The map p is a local homeomorphism.* Take any point $x \in X$ and denote by $p|_U \colon (U, x) \to U$ the restriction of p to any basis neighborhood (U, x) of x, so that U will be an open path connected and simply connected set in B. The path connectedness of U implies the surjectivity of $p|_U$ and its simple connectedness, the injectivity of $p|_U$.

(iii) *The subgroup $p_\#(\pi_1(X))$ coincides with G.* Let α be a loop in B with basepoint b_0 and $\widetilde{\alpha}$ be the lift of α initiating at x_0 ($\widetilde{\alpha}$ is not necessarily a closed path). The subgroup $p_\#(\pi_1)(X)$ consists of homotopy classes of the loops α whose lifts $\widetilde{\alpha}$ are closed paths. By construction, the path $\widetilde{\alpha}$ is closed iff the equivalence class of the loop α corresponds to the point x_0, i.e., if the homotopy class of α is an element of G.

(iv) *The map p is unique.* We omit the proof of this fact here. □

11.4. Regular coverings

A covering $p\colon T \to B$ is called *regular* if the subgroup $p_\#(\pi_1(E)) \subset \pi_1(B)$ is normal.

Theorem 11.4. *If $p\colon E \to B$ is any regular covering, then the quotient group $\pi(B)/p_\#(\pi_1(E))$ is isomorphic to the group of deck transformations of the fiber $D = p^{-1}(b_0)$.*

Proof. There is a natural bijection between the right cosets ("правые смежные классы" in Russian) of the subgroup $p_\#(\pi_1(E)) \subset \pi_1(B)$ and D, but since this subgroup is normal, these cosets forms a group that "shuffles" the points of D, so that the quotient group $\pi(B)/p_\#(\pi_1(E))$ is the group of deck transformations of D. □

11.5. Exercises

11.1. Suppose that one surface is covered by another surface. What is the relation between their Euler characteristics, if the covering is n-fold?

11.2. Prove that the sphere with g_1 handles can be covered by the sphere with g_2 handles ($g_1, g_2 \geqslant 2$) iff $g_1 - 1$ is a divisor of $g_2 - 1$.

11.3. Construct a nonregular covering of the wedge product of two circles.

11.4. Construct two regular coverings of the wedge product of two circles that are not homotopy equivalent to each other.

11.5. Prove that for any $n \geqslant 2$ the wedge product of two circles can be covered by the wedge product of n circles.

11.6. Prove that if the base surface of a covering $p\colon N^2 \to M^2$ is orientable, then so is the covering surface N^2.

11.7. Let X be the union of the lateral surface of the cone and the half-line issuing from its vertex v, and let $p\colon X \to B$ be the natural projection of X on the line $B = \mathbb{R}$. Show that $p\colon X \to B$ does not possess the covering homotopy property.

11.8. Let the covering surface N^2 of a covering $p\colon N^2 \to M^2$ is orientable. Is it true that the base surface M^2 is orientable?

11.9. Can $\mathbb{R}P^2$ cover the sphere?

11.10. Can the torus $\mathbb{T}^2$ cover $\mathbb{T}^2$ by a 3-fold covering?

11.11. Can $\mathbb{R}P^2$ be covered by the plane?

11.12. Construct the universal covering of the Möbius band.

11.13. Construct the universal covering of the torus $\mathbb{T}^3$.

11.14. Construct the universal covering of the wedge of two circles.

11.15. Construct the universal covering of the wedge $\mathbb{S}^1 \vee \mathbb{S}^2$.

11.16. Construct the universal covering of the sphere with $g \geqslant 2$ handles.

11.17. Suppose some connected graph G has e edges and v vertices. Find the fundamental group of the graph G.

11.18. Prove that any subgroup of a free group is a free group.

11.19. Prove that the free group of rank 2 contains as a subgroup the free group of rank n for all n (including $n = \infty$).

11.20. Give an example of a covering space $p\colon E \to B$, of a space A, and a map $f\colon A \to B$ that cannot be lifted to E.

11.21. Prove that the universal cover $\omega\colon U \to B$ of any (pathconnected) space B is the cover of any other covering of B, i.e., for any covering space $p\colon E \to B$, there exists a covering space $q\colon U \to E$.

Lecture 12

Knots and links

Knot theory, which studies knots, links ("зацепления" in Russian), and their invariants, has a long history that begins at the end of the 18th century (Vandermonde), with significant contributions by Gauss, Poincaré, Reidemeister, Alexander, Conway, Fox. The theory flourished at the end of the 20th century. Four Fields medalists—Jones, Witten, Drinfeld, Kontsevich—worked in knot theory; other leading researchers are Kauffman, Reshetikhin, Turaev, Viro, Vassiliev, Khovanov. The theory is still going strong today.

12.1. Main definitions

A *knot* is a closed non-self-intersecting broken line in $\mathbb{R}^3$, a *link* is a set of nonintersecting and non-self-intersecting broken lines in $\mathbb{R}^3$. Two knots or links are called *isotopic* if there exists a finite sequence of Δ-moves (see Figure 12.1) transforming one into the other.

Note that in the definition of Δ-moves the knot (link) does not intersect the triangle ABC defining the move, except along its sides, as shown in the figure. In the definition of Δ-moves, we include the case when the

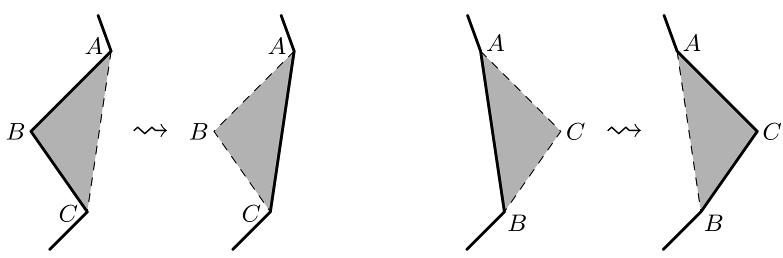


Figure 12.1. Δ-moves

triangle ABC is degenerate; in that case the move reduces to adding or removing a vertex.

An example of a sequence of Δ-moves changing the shape of a knot are shown in Figure 12.2. The figure clarifies the idea that two knots are isotopic if their practical models (made from rope) can be given the same shape by appropriately moving the ropes.

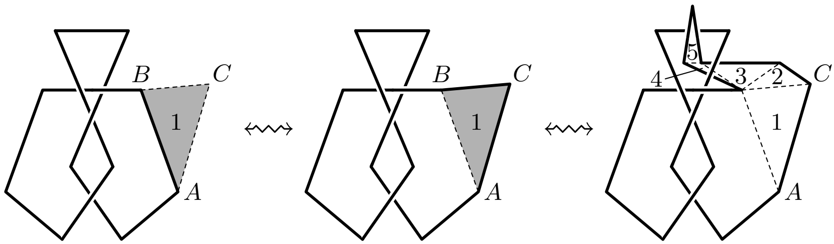


FIGURE 12.2. A sequence of Δ-moves

Isotopy is obviously an equivalence relation, and the word "knot" is often used in the sense of "isotopy class of knots", we often say that two closed broken lines are "the same knot" or "have the same knot type" if they are isotopic. (And similarly for links.)

A knot is called *trivial* (or said to be the *unknot*) if it is isotopic to a regular polygon. Examples of famous knots and links are shown in Figure 12.2. In the figure the knots are presented in the form of *knot*

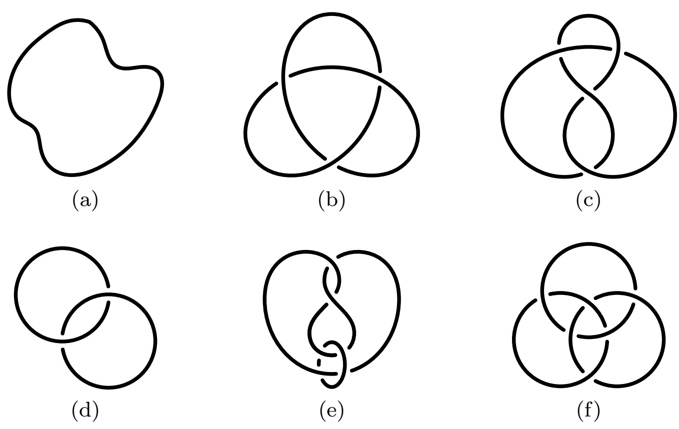


FIGURE 12.3. (a) unknot; (b) right trefoil; (c) eight knot; (d) Hopf link; (e) Whitehead link; (f) Borromeo rings

diagrams, i.e., projections of the knot on the plane in general position, with underpass-overpass information at each double point (it shows which one of the branches lies above the other).

There are several equivalent definitions of knot, link, and isotopy (e. g. smooth knots or PL-knots). The definition given above is the most elementary one. Another elementary definition of knot consists in "putting each knot in a box", i.e., defining a knot as a broken line inside a cube joining the centers of opposite faces of the cube, with isotopy being a sequence of Δ-moves performed inside the cube and not moving the endpoints of the broken line. It is easy to show that there is natural bijection between isotopy classes of boxed knots and knot types as defined above.

12.2. The arithmetic of knots

We define the *composition* (also called *connected sum*) of two (boxed) knots K_1, K_2 as the knot $K_1\#K_2$ obtained by fitting the boxes together as shown in Figure 12.4). Under the composition operation *knots form a semigroup* denoted by $\mathcal{K}$.

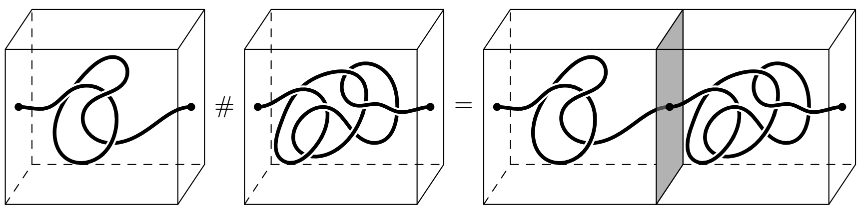


FIGURE 12.4. Composition of two knots

A knot K is called *prime* if it cannot be presented as the sum of two nontrivial knots, i.e., $K = K_1\#K_2$ implies that either K_1 or K_2 is the unknot.

Theorem 12.1. *The semigroup $\mathcal{K}$ is commutative, it has no inverse elements (i.e., $K_1\#K_2 = \bigcirc$, where $\bigcirc$ denotes the unknot, implies that both K_1 and K_2 are trivial) and each nontrivial knot possesses a unique (up to order) decomposition into prime knots.*

We omit the (fairly difficult) proof of this lovely theorem, but show the isotopy demonstrating commutativity (Figure 12.5.)

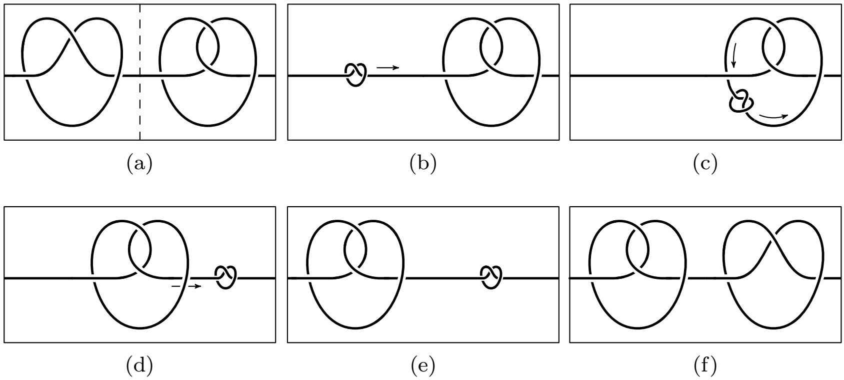


FIGURE 12.5. Commutativity of the connected sum of knots

12.3. The combinatorics of knots: Reidemeister moves

The knot classification problem is a very difficult three-dimensional geometric problem (for a detailed formulation, see Section 12.5 below), but it has been reduced to a combinatorial two-dimensional problem by Reidemeister. This reduction was done by means of certain modifications of knot diagrams called *Reidemeister moves*; they are shown in Figure 12.6. The first move, Ω_1, is the removal (addition) of a small loop, the second one, Ω_2, is the removal (addition) of an overlap, and the third, Ω_3, is the passage of a branch of the knot over a crossing point.

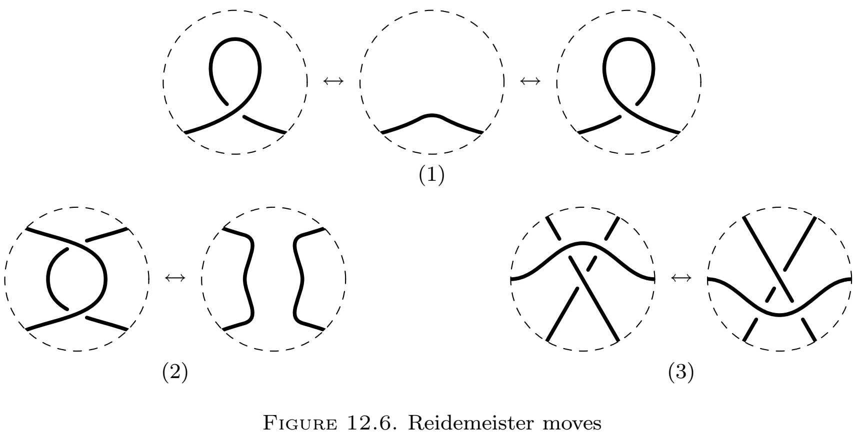


FIGURE 12.6. Reidemeister moves

Theorem 12.2. *If two knot* (*link*) *diagrams define isotopic knots* (*links*), *then one can be taken to the other by a finite sequence of Reidemeister moves.*

We omit the proof, which can be obtained by a general position argument.

The Reidemeister theorem did not lead to a simple solution of the knot classification problem, but turned out to be extremely useful in the construction of various knot invariants.

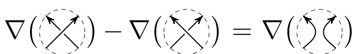
12.4. The Alexander–Conway polynomial

The Alexander–Conway polynomial is an invariant of oriented knots and links; it can be introduced by means of the Conway axioms: we are given a rule that to each oriented diagram of a knot or link L assigns a polynomial $\nabla_L(x)$ and satisfies the following axioms:

I. [Invariance] $\nabla_L(x)$ *is an isotopy invariant.*

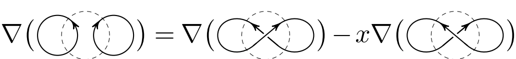
II. [Normalization] *For the unknot* $\bigcirc$, $\nabla_{\bigcirc}(x) = 1$.

III. [Skein relation] *The following equality holds*:



for any three link diagrams that are identical everywhere except inside the dotted circles, where they are as shown in the figure.

We will not prove that the Alexander–Conway polynomial exists and is well defined by these axioms, we only present an example of its calculation.



$= \nabla(\bigcirc) - \nabla(\bigcirc) = 1 - 1 = 0.$

12.5. About the classification of knots

The solution of the *knot classification problem* is an algorithm that determines whether or not two knot diagrams define the same knot. The existence of such an algorithm was proved by S. V. Matveev a few years ago, but the algorithm is too complicated to be implemented in a computer. However, prime knots with a small ($\leqslant 16$) number of crossings have been classified (by means of invariants more powerful than the Alexander–Conway polynomial) and are tabulated in *knot tables.* A small knot table is given below.

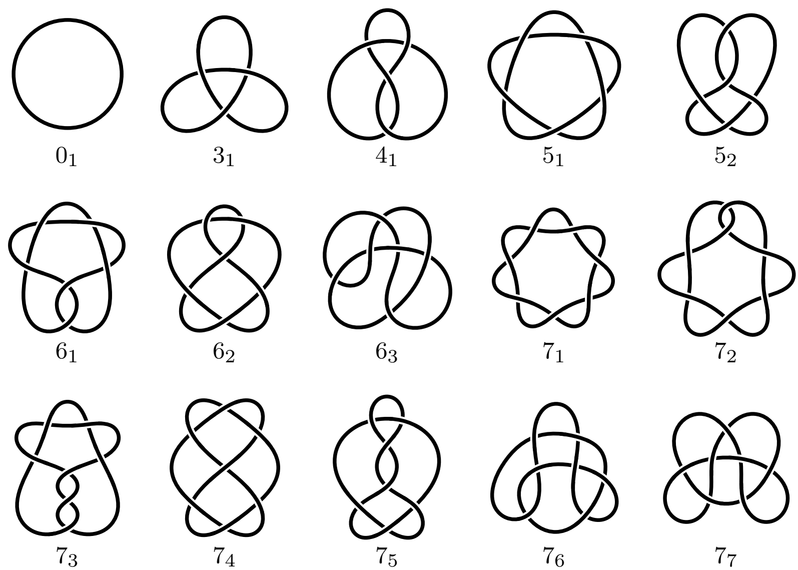


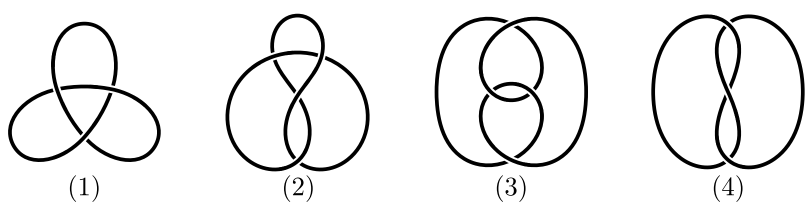
FIGURE 12.7. Table of prime knots with 7 crossings or less

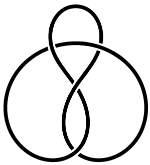
A particular case of the knot classification problem is the *unknotting problem*: to find an algorithm that decides whether a given knot diagram is the unknot. Such an algorithm has been constructed by Ivan Dynnikov, it is known that it gives the correct answer for knot diagrams with 500 crossings or less, but it has not been proved that it terminates with the correct answer when there are more than 500 crossings.

These topics are the subject of ongoing research.

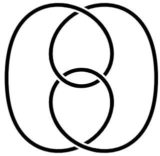
12.6. Exercises

12.1. Which of the knots in the picture are trivial, are trefoils, are eight knots?

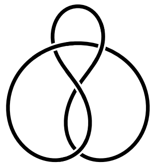




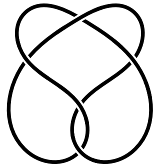
(5)



(6)



(7)



(8)

12.2. Compute the Conway polynomials of the two Hopf links.

12.3. Compute the Conway polynomials of the two (right and left) trefoils.

12.4. Compute the Conway polynomials of the eight knot (4_1 in the knot table).

Contents