Topology I—III, HU Berlin

Chris Wendl

Contents

Preface

The original version of these notes was created in 2018–19 for a two-semester sequence of topology courses at Humboldt University, Berlin. It has undergone substantial revisions during a few repetitions of the two-semester course since then, plus the addition of a third semester in 2025. The topics are divided among the three semesters roughly as follows:

- First semester: basic point-set topology, fundamental group and covering spaces, manifolds of dimension one and two, introduction to homology
- Second semester: homology and cohomology
- Third semester: homotopy theory, higher homotopy groups, fiber bundles and characteristic classes

A few topics appear in multiple semesters, e.g. while the end of the first semester contains material on singular homology, the second semester does not assume previous knowledge of homology, and thus starts that subject from the beginning, though at a slightly higher level of sophistication. This reflects the fact that at our university, Topology I is technically a Bachelor-level course and Topology II is technically Master-level, though in practice, the audience for both courses is typically a mixture.

There is a nearly exact one-to-one correspondence between the chapters in these notes and the actual 90-minute lectures given in the course, though for some chapters that are a bit fatter, some portions had to be skipped or mentioned only briefly in class.

Since the notes were designed for use at a German university, I have made an effort to include the German translations (*geschrieben in dieser Schriftart*) of important terms wherever they are introduced. The reader may notice that this effort subsides later in the course, as the deeper one gets into algebraic topology, the harder it becomes to find authoritative German sources for clarifying the terminology (and I am not linguistically qualified to invent terms in German myself).

About the current version

The version you are looking at right now is being updated regularly in order to serve as lecture notes for the HU's Topology II course in the Winter 2024–25 semester, and it is intended to continue into the Summer 2025 semester as lecture notes for Topology III. I did not teach the Topology I course that immediately preceded those two semesters, but my lecture notes nonetheless closely resemble the course that was actually taught.

One innovation of the current version—implemented in the notes for the second semester but not yet for the first semester—is that all exercises now appear in their own subsection at the end of each lecture, and some of them are marked with an asterisk (like this (*****)). The asterisk means that the exercise is *essential*, e.g. because it contains a proof of some important result that will be used again in the course, perhaps multiple times. Exercises without an asterisk are intended to be helpful and/or informative, but not essential for the logical continuity of the notes.

Most recent update: **November 22, 2024**

Disclaimer and acknowledgements

These lecture notes were written quickly, and while many typos have in the mean time been eliminated due to careful reading by a few motivated students, some probably remain. If you notice any, please send me an e-mail and I will correct. Thanks for corrections already received are due to Lennard Henze, Jens Lücke, Mateusz Majchrzak, Marie Christin Schmidtlein, Rens Breur, Maxim Nevkrytyh, Laurenz Upmeier zu Belzen, Florian Kaufmann, Ben Eltschig and Daniel Acker. (Apologies if I forgot anyone!)

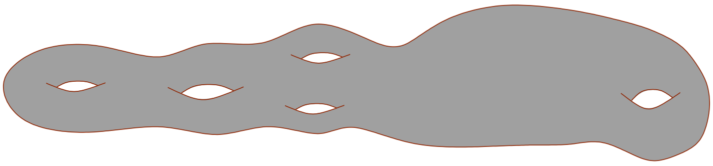
First semester (Topologie I)

1. Introduction and motivation

To start with, let us discuss what kinds of problems are studied in topology. This lecture is only intended as a sketch of ideas, so nothing in it is intended to be precise—we'll introduce precise definitions in the next lecture.

(1) Classification of spaces. Let's assume for the moment that we understand what the word "space" means. We'll be more precise about it next week, but in this course, a "space" X is a set with some extra structure on it such that we have well-defined notions of things like *open* subsets (*offene Teilmengen*) $\mathcal{U} \subset X$ and *continuous maps/mappings* (*stetige Abbildungen*) $f : X \to Y$ (where Y is another space). It is then natural to consider two spaces X and Y equivalent if there is a **homeomorphism** (*Homöomorphismus*) between them: this means a continuous bijection $f : X \to Y$ whose inverse $f^{-1} : Y \to X$ is also continuous. We say in this case that X and Y are **homeomorphic** (*homöomorph*).

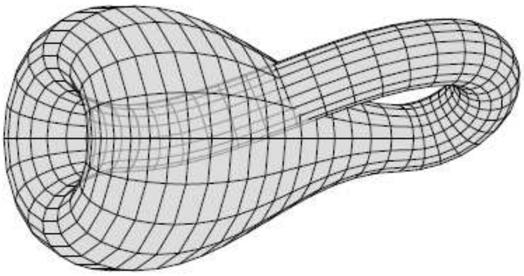
So for instance, one can try to classify all *surfaces* (*Flächen*) up to homeomorphism:



The space in this picture is known as a "closed orientable surface of genus (*Geschlecht*) five". The genus is a nonnegative integer that, roughly speaking, counts the number of "handles" you would need to attach to a sphere in order to construct the surface. The notation Σ_g is often used for a surface of genus $g \geqslant 0$.

There are also closed surfaces that cannot be embedded in $\mathbb{R}^3$, though they are harder to visualize. Here are two examples.

Example 1.1. Here is a picture of the **Klein bottle** (*Kleinsche Flasche*), a surface that can be "immersed" (with self-intersections) in $\mathbb{R}^3$, but not embedded:



We'll give a more precise definition of the Klein bottle as a topological space later.

EXAMPLE 1.2. The **real projective plane** (*reelle projektive Ebene*) $\mathbb{RP}^2$ is a space that can be described in various equivalent ways:

(1) $\mathbb{RP}^2 := S^2/\!\sim$, i.e. the set of equivalence classes of elements in the unit sphere $S^2 := \{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| = 1\}$, with the equivalence relation defined by $\mathbf{x} \sim -\mathbf{x}$ for each $\mathbf{x} \in S^2$. In other words, every element of $\mathbb{RP}^2$ is a set of two elements $\{\mathbf{x}, -\mathbf{x}\}$, with both belonging to the unit sphere. (See Remark 1.3 below on notation for defining equivalence relations.)
(2) $\mathbb{RP}^2 := \mathbb{D}^2/\!\sim$, where $\mathbb{D}^2 := \{\mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x}| \leqslant 1\}$ and the equivalence relation is defined by $z \sim -z$ for every point z on the boundary of the disk. One obtains this from the first description of $\mathbb{RP}^2$ by restricting attention to only one hemisphere of S^2; no information is lost since the other hemisphere is identified with it, but along the equator between them, there is still an identification of antipodal points.
(3) $\mathbb{RP}^2$ is the space of all lines through 0 in $\mathbb{R}^3$. This is equivalent to the first description since every line through the origin in $\mathbb{R}^3$ hits S^2 at exactly two points, which are antipodal to each other.
(4) $\mathbb{RP}^2$ is the space constructed by gluing a disk $\mathbb{D}^2$ to a **Möbius strip** (*Möbiusband*)
$$\mathbb{M} := \left\{(\theta, t\cos(\pi\theta), t\sin(\pi\theta)) \in \mathbb{R}/\mathbb{Z} \times \mathbb{R}^2 \;\middle|\; \theta \in \mathbb{R},\ t \in [-1,1]\right\}.$$
To see this, draw a picture of the unit sphere S^2 and think of $\mathbb{RP}^2$ as $S^2/\!\sim$. After identifying antipodal points of the sphere in this way, a neighborhood of the equator looks like a Möbius strip, and everything else is a disk (it looks like two disks in the picture, but the two are identified with each other).

More generally, for each integer $n \geqslant 0$ one can define the n-**sphere**
$$S^n = \left\{\mathbf{x} \in \mathbb{R}^{n+1} \;\middle|\; |\mathbf{x}| = 1\right\}$$
and the **real projective** n-**space**
$$\mathbb{RP}^n = S^n/\{\mathbf{x} \sim -\mathbf{x}\} = \left\{\text{lines through } 0 \text{ in } \mathbb{R}^{n+1}\right\}.$$

REMARK 1.3. In topology, we often specify an equivalence relation $\sim$ on a set X with words such as "the equivalence relation defined by $x \sim f(x)$ for all $x \in A$" where $A \subset X$ is a subset and $f : A \to X$ a map. This should always be interpreted to mean that $\sim$ is the *smallest* equivalence relation for which the stated property is true, i.e. since every equivalence relation must also be reflexive and symmetric, it is implied that $x \sim x$ for all $x \in X$ and $f(x) \sim x$ for all $x \in A$, even if we do not say so explicitly. Transitivity may then imply further equivalences that are not explicitly specified: for an extreme example, "the equivalence relation on $\mathbb{Z}$ such that $n \sim n+1$ for all $n \in \mathbb{Z}$" makes every integer equivalent to every other integer, i.e. there is only one equivalence class.

Here is a result we will be able to prove later in the course:

THEOREM 1.4. *A closed orientable surface Σ_g of genus g is homeomorphic to a closed orientable surface Σ_h of genus h if and only if $g = h$.*

The hard part is showing that if $g \neq h$, then there cannot exist any continuous bijective map $f : \Sigma_g \to \Sigma_h$ with a continuous inverse. This requires techniques from the subject known as *algebraic* topology. The main idea will be that we can associate to each topological space X an algebraic object (e.g. a group) $H(X)$ such that any continuous map $f : X \to Y$ induces a homomorphism $f_* : H(X) \to H(Y)$, and such that compositions of continuous maps satisfy
$$(f \circ g)_* = f_* \circ g_*$$
and the identity map $\mathrm{Id} : X \to X$ gives rise to the identity map $H(X) \to H(X)$. These properties imply that whenever $f : X \to Y$ is a homeomorphism, $f_* : H(X) \to H(Y)$ must be an

isomorphism. Thus it suffices to compute the algebraic objects $H(\Sigma_g)$ and $H(\Sigma_h)$ and show that they are not isomorphic. (Recognizing non-isomorphic groups is often easier than recognizing non-homeomorphic spaces.)

The full classification of closed orientable surfaces up to homeomorphism is completed by the following result:

Theorem 1.5. *Every closed connected and orientable surface is homeomorphic to* Σ_g *for some* $g \geqslant 0$.

The previous theorem implies of course that for any given surface, the value of g in this result is unique. For the moment, you can understand the word "orientable" to mean "embeddable in $\mathbb{R}^3$". There is a similar result for the non-orientable surfaces: notice that by the fourth definition we gave above for $\mathbb{RP}^2$, one can understand $\mathbb{RP}^2$ as the result of taking S^2, cutting out a hole (e.g. removing the southern hemisphere, thus leaving the northern hemisphere, which is also a disk $\mathbb{D}^2$) and then gluing in a Möbius strip. That is the first example of the following more general construction:

Theorem 1.6. *Every closed connected and non-orientable surface is homeomorphic to a surface obtained from* S^2 *by cutting out finitely many holes and gluing in Möbius strips.*

Surfaces are the simplest interesting examples of more general topological spaces called **manifolds** (*Mannigfaltigkeiten*): a surface is a 2-dimensional manifold, while a smooth curve such as the circle S^1 is a 1-dimensional manifold. In general, one can consider n-dimensional manifolds (abbreviated as "n-manifolds") for any integer $n \geqslant 0$; obvious examples include $\mathbb{R}^n$, S^n and $\mathbb{RP}^n$. The classification problem becomes much harder when $n \geqslant 3$, e.g. the following difficult problem was open for almost exactly 100 years:

Poincaré conjecture (solved by G. Perelman, c. 2004). *Suppose* X *is a closed and connected* 3*-manifold that is "simply connected" (i.e. every continuous map* $f : S^1 \to X$ *can be extended continuously to* $\mathbb{D}^2 \to X$*). Then* X *is homeomorphic to* S^3.

One of the more surprising developments in topology in the 20th century was that the analogue of this problem in dimensions greater than three turns out to be easier. We'll introduce the notion of "homotopy equvalence" (*Homotopieäquivalenz*) in a few weeks; it turns out that for closed 3-manifolds, the condition of being *simply connected* is equivalent to being homotopy equivalent to S^3. Thus the following two results are higher-dimensional versions of the Poincaré conjecture, but they were proved much earlier:

Theorem 1.7 (S. Smale, c. 1960). *For every* $n \geqslant 5$, *every closed connected* n*-manifold homotopy equivalent to* S^n *is also homeomorphic to* S^n.

Theorem 1.8 (M. Freedman, c. 1980). *Every closed connected* 4*-manifold homotopy equivalent to* S^4 *is also homeomorphic to* S^4.

(2) Differential topology. Though we will not have much time to talk about it in this semester, the neighboring field of "differential" topology modifies the classification problem by studying the following stronger notion of equivalence between spaces: X and Y are **diffeomorphic** (*diffeomorph*) if there exists a homeomorphism $f : X \to Y$ such that both f and f^{-1} are infinitely differentiable, i.e. C^∞, and f is in this case called a **diffeomorphism** (*Diffeomorphismus*). From your analysis courses, you at least know what this means if X and Y are open subsets of Euclidean spaces—defining "differentiability" on spaces more general than that requires some notions from the subject of *differential geometry.* In a nutshell, it requires X and Y to be spaces on which any map $X \to Y$ can at least *locally* (i.e. in a sufficiently small neighborhood of any point) be identified with a map between open subsets of Euclidean spaces, for which we know how to define derivatives.

Identifying a small neighborhood in X with an open subset of $\mathbb{R}^n$ is another way of saying that we can choose a set of n independent "coordinates" to describe the points in that neighborhood, and this is the fundamental property that defines X as an n-dimensional manifold. So talking about smooth maps and diffeomorphisms doesn't make sense for arbitrary topological spaces, but it does make sense for at least some class of manifolds, and these are the main objects of study in differential topology.

It turns out that up to dimension three, classification up to diffeomorphism is equivalent to classification up to homeomorphism:

THEOREM 1.9. *For $n \leqslant 3$, two n-manifolds X and Y are diffeomorphic if and only if they are homeomorphic.*

For $n = 1$ and $n = 2$, this theorem can be explained by the fact that both versions of the classification problem for n-manifolds are not that hard to solve explicitly (this was already understood in the 19th century), and the answer for both versions turns out to be the same. The story of $n = 3$ is much more complicated, as a complete classification of 3-manifolds is not known, but this theorem was proved in the first half of the 20th century by using the more combinatorial notion of "piecewise linear" manifolds as an intermediary notion between "smooth" and "topological" manifolds.

From dimension four upwards, all hell breaks loose. For example, there are "exotic" $\mathbb{R}^4$'s:

THEOREM 1.10. *There exist 4-manifolds that are homeomorphic but not diffeomorphic to $\mathbb{R}^4$.*

And from dimension seven upwards, there also tend to exist "exotic spheres":

THEOREM 1.11 (Kervaire and Milnor, 1963). *There exist exactly 28 distinct manifolds that are homeomorphic to S^7 but not diffeomorphic to each other.*

As you might guess, there is an algebraic phenomenon behind the appearance of the number 28 in this theorem: it is the order of a group. In every dimension n, one can define a group structure on the set of all smooth manifolds up to diffeomorphism that are homeomorphic to S^n. Milnor and Kervaire proved that when $n = 7$, this group has order 28. In the mean time, this group is quite well understood in most cases: it is sometimes trivial (e.g. for $n = 1, 2, 3, 5, 6$) and often nontrivial, but always finite. The only case for which almost nothing is known is $n = 4$; dimension four turns out to be the hardest case in differential topology, because it is on the borderline between "low dimensional" and "high dimensional" methods, where often neither set of methods applies. If you can solve the following open problem, you deserve an instant Ph.D. (and also a permanent job as a research mathematician, and possibly a Fields medal):

CONJECTURE 1.12 ("smooth Poincaré conjecture"). *Every manifold homeomorphic to S^4 is also diffeomorphic to S^4.*

It is difficult to say whether this conjecture is generally believed to be true or false.

(3) Fixed point problems. Here is a simpler class of problems on which we'll actually be able to prove something in this semester. Suppose $f : X \to X$ is a continuous map. We say $x \in X$ is a **fixed point** (*Fixpunkt*) of f if $f(x) = x$. The question is: under what assumptions on X is f guaranteed to have a fixed point? Note that this is fundamentally different from the fixed point results you've probably seen in analysis, e.g. the Banach fixed point theorem (also known as the *contraction mapping principle*) is a result about a special class of maps satisfying analytical conditions, it does not just apply to *every* continuous map on a certain space.

The simplest fixed point theorem in topology is a statement about maps on the n-dimensional disk $\mathbb{D}^n := \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x}| \leqslant 1\}$.

Theorem 1.13 (Brouwer's fixed point theorem). *For every integer $n \geqslant 1$, every continuous map $f : \mathbb{D}^n \to \mathbb{D}^n$ has a fixed point.*

The case $n = 1$ is an easy consequence of the intermediate value theorem, but for $n \geqslant 2$, we need some techniques from algebraic topology. Here is a sketch of the argument; we will fill in the gaps over the course of the semester.

We argue by contradiction, so suppose there exists a continuous map $f : \mathbb{D}^n \to \mathbb{D}^n$ such that $f(x) \neq x$ for every $x \in \mathbb{D}^n$. Then there is a unique line in $\mathbb{R}^n$ connecting $f(x)$ to x for each $x \in \mathbb{D}^n$. Let $g(x) \in S^{n-1}$ denote the point on the boundary of $\mathbb{D}^n$ obtained by following the unique line from $f(x)$ through x until that line reaches the boundary of the disk. Note that if x is already on the boundary, then by this definition $g(x) = x$. It is not hard to convince yourself that what we've just defined is a continuous map

$$g : \mathbb{D}^n \to S^{n-1},$$

and if $i : S^{n-1} \hookrightarrow \mathbb{D}^n$ denotes the natural inclusion map for the subset $S^{n-1} \subset \mathbb{D}^n$, then g satisfies

$$g \circ i = \mathrm{Id}_{S^{n-1}} . \tag{1.1}$$

We claim that, actually, no such map can exist. The proof of this requires an algebraic invariant, whose complete construction will require some time and effort, but for now I'll just tell you the result: one can associate to each space X an abelian group $H_{n-1}(X)$ called the **singular homology** (*singuläre Homologie*) of X in dimension $n - 1$, which satisfies the usual desirable properties that continuous maps $f : X \to Y$ induce group homomorphisms $f_* : H_{n-1}(X) \to H_{n-1}(Y)$ satisfying $(f \circ g)_* = f_* \circ g_*$ and $\mathrm{Id}_* = \mathbb{1}$. Crucially, one can also compute this invariant for both $\mathbb{D}^n$ and S^{n-1}, and the answers are

$$H_{n-1}(\mathbb{D}^n) = \{0\}, \qquad H_{n-1}(S^{n-1}) \cong \mathbb{Z}.$$

Now the relation (1.1) implies that $g_* \circ i_*$ is the identity map on $H_{n-1}(S^{n-1}) \cong \mathbb{Z}$, so in particular it is an isomorphism. But $g_* \circ i_*$ also factors through the trivial group $H_{n-1}(\mathbb{D}^n) \cong \{0\}$, and therefore can only be the trivial homomorphism. This is a contradiction, thus proving Brouwer's theorem.

We will discuss the construction of singular homology and carry out the required computations for the above argument in the last few weeks of this semester; homology and the closely related subject of **cohomology** (*Kohomologie*) will then be the main topic of Topology 2 next semester. But before all that, we will also spend considerable time on other invariants in algebraic topology, notably the fundamental group, which underlies the notion of "simply connected" spaces appearing in the Poincaré conjecture.

2. Metric spaces

We now begin in earnest with point-set topology, which will be the main topic for the next three or four weeks. This subject is important but a little dry, so we will cover only the portions of it that seem absolutely necessary as groundwork for studying the more geometrically motivated questions discussed in the previous lecture.

The subject begins with metric spaces, because these are the most familiar examples of topological spaces. For most students, this material will be a review of things you've seen before in analysis courses. Almost everything in this lecture will be generalized to a wider and slightly more abstract context when we introduce topologies and topological spaces next week.

Definition 2.1. A **metric space** (*metrischer Raum*) is a set X endowed with a function $d : X \times X \to \mathbb{R}$ that satisfies the following conditions for all $x, y, z \in X$:

(i) $d(x, y) \geqslant 0$;

(ii) $d(x,x) = 0$;
(iii) $d(x,y) = d(y,x)$, i.e. "symmetry";
(iv) $d(x,z) \leqslant d(x,y) + d(y,z)$, i.e. the "triangle inequality" (*Dreiecksungleichung*);
(v) $d(x,y) > 0$ whenever $x \neq y$.

The function d is then called a **metric** (*Metrik*). If d satisfies the first four conditions but not necessarily the fifth, then it is called a **pseudometric** (*Pseudometrik*).

Much of the theory of metric spaces makes sense for pseudometrics just as well as metrics, but we will see that some desirable and intuitively "obvious" facts become false when the positivity condition is dropped.

In any metric space (X,d), one can define the **open ball** (*offene Kugel*) of radius $r > 0$ about a given point $x \in X$ as

$$B_r(x) := \left\{ y \in X \;\middle|\; d(x,y) < r \right\}.$$

An arbitrary subset $\mathcal{U} \subset X$ is then called **open** (*offen*) if for every $x \in \mathcal{U}$, the ball $B_\epsilon(x)$ is contained in $\mathcal{U}$ for all $\epsilon > 0$ sufficiently small. (Of course it only needs to be true for one particular $\epsilon > 0$, since then it is true for all smaller ϵ as well.) Given a subset $A \subset X$, another subset $\mathcal{U} \subset X$ is called a **neighborhood** (*Umgebung*) of A in X if $\mathcal{U}$ contains some open subset of X that also contains A. Some books require the neighborhood itself to be open, but we will not require this; it makes very little difference in practice, but this bit of extra freedom in our definition will allow us to make certain other definitions and proofs a few words shorter now and then.

A subset $A \subset X$ is **closed** (*abgeschlossen*) if its complement $X \backslash A$ is open. Achtung: this is not the same thing as saying that A is not open. It is a common trap for beginners to think that every subset must be either open or closed, but in reality, most are neither—and some (e.g. X itself) are both.[1]

Whenever you encounter a set of axioms, you should ask yourself why we are studying these axioms in particular—why not a slightly different set of axioms? In the case of metrics, it's fairly obvious why we would want any notion of "distance" to satisfy conditions (i)–(iii) and (v), but perhaps the triangle inequality seems slightly less obvious. So, let us point out two obviously desirable properties that follow mainly from the triangle inequality:

- The "open ball" $B_r(x) \subset X$ is also an open subset in the sense of the definition given above. Indeed, for any $y \in B_r(x)$, we have $B_\epsilon(y) \subset B_r(x)$ for every $\epsilon < r - d(x,y)$ since every $z \in B_\epsilon(y)$ then satisfies
$$d(x,z) \leqslant d(x,y) + d(y,z) < d(x,y) + \epsilon < d(x,y) + r - d(x,y) = r.$$
- The function $d : X \times X \to [0,\infty)$ is *continuous* (see below for a review of the definition of continuity), since one can use the triangle inequality to show that for every $x, y, x', y' \in X$,
$$|d(x,y) - d(x',y')| \leqslant d(x,x') + d(y,y').$$

Also, while I'm sure you already accept without question that the distance between two distinct points should always be positive rather than zero, let us point out one "obvious" fact that would cease to be true if condition (v) were removed:

- For every $x \in X$, the subset $\{x\} \subset X$ is closed. Indeed, $X \backslash \{x\}$ is an open subset of X because for every $y \in X \backslash \{x\}$, the ball $B_\epsilon(y)$ is contained in $X \backslash \{x\}$ for all $\epsilon < d(x,y)$. (This of course presupposes that $d(x,y) > 0$.)

You're probably not used to thinking about pseudometric spaces much, so here is an example.

---

[1] Yes, the empty set $\varnothing \subset X$ is always open. Reread the definition carefully until you are convinced that this is true.

EXAMPLE 2.2. Let $X = (\mathbb{R} \times \{0,1\})/\sim$ for an equivalence relation defined by $(x,0) \sim (x,1)$ for every $x \neq 0$. We can think of this intuitively as a "real line with two zeroes" because it mostly looks just the same as $\mathbb{R}$ (each number $x \neq 0$ corresponding to the equivalence class of $(x,0)$ and $(x,1)$), but $x = 0$ is an exception, where there really are *two* distinct points $[(0,0)]$ and $[(0,1)]$ in X. We can then define $d : X \times X \to \mathbb{R}$ by

$$d([(x,i)],[(y,j)]) := |x-y| \quad \text{for } i,j \in \{0,1\},\ x,y \in \mathbb{R}.$$

This satisfies conditions (i)–(iv) for all the same reasons that the usual metric on $\mathbb{R}$ does, but condition (v) fails because

$$d([(0,0)],[(0,1)]) = 0$$

even though $[(0,0)] \neq [(0,1)]$.

EXERCISE 2.3. Show that for the pseudometric space X in Example 2.2, $\{[(0,0)]\} \subset X$ is not a closed subset.

DEFINITION 2.4. In a metric space (X,d), a sequence (*Folge*) $x_n \in X$ indexed by $n \in \mathbb{N}$ **converges to** (*konvergiert gegen*) a point $x \in X$ if for every $\epsilon > 0$, we have $x_n \in B_\epsilon(x)$ for all n sufficiently large. Equivalently, this means that for every neighborhood $\mathcal{U} \subset X$ of x, $x_n \in \mathcal{U}$ for all n sufficiently large. We use the notation

$$x_n \to x \qquad \text{or} \qquad \lim x_n = x$$

to indicate that x_n converges to x.

Note that in the second formulation of this definition, involving arbitrary neighborhoods instead of the open ball $B_\epsilon(x)$, one can understand the definition without knowing what the metric is—one only has to know what a "neighborhood" is, which means knowing which subsets are open and which are not. This will be the formulation that we need when we generalize sequences and convergence to arbitrary topological spaces.

Here is a similarly standard definition from analysis, for which we give three equivalent formulations.

DEFINITION 2.5. For two metric spaces (X,d_X) and (Y,d_Y), a map (*Abbildung*) $f : X \to Y$ is called **continuous** (*stetig*) if it satisfies any of the following equivalent conditions:

(a) For every $x_0 \in X$ and $\epsilon > 0$, there exists a number $\delta > 0$ such that $d_Y(f(x),f(x_0)) < \epsilon$ whenever $d_X(x,x_0) < \delta$, i.e. $f(B_\delta(x_0)) \subset B_\epsilon(f(x_0))$.

(b) For every open subset $\mathcal{U} \subset Y$, the preimage

$$f^{-1}(\mathcal{U}) := \{x \in X \mid f(x) \in \mathcal{U}\}$$

is an open subset of X.

(c) For every convergent sequence $x_n \in X$, $x_n \to x$ implies $f(x_n) \to f(x)$.

The equivalence of (a) and (b) is pretty easy to see: if (a) holds and $\mathcal{U} \subset Y$ is open, then for every $x_0 \in f^{-1}(\mathcal{U})$, the openness of $\mathcal{U}$ guarantees an $\epsilon > 0$ such that $f(x_0) \in B_\epsilon(f(x_0)) \subset \mathcal{U}$. But then condition (a) gives a $\delta > 0$ such that $f(B_\delta(x_0)) \subset B_\epsilon(f(x_0)) \subset \mathcal{U}$, implying $B_\delta(x_0) \subset f^{-1}(\mathcal{U})$, hence $\mathcal{U}$ is open and (b) therefore holds. Conversely, if (b) holds, then (a) holds because $B_\epsilon(f(x_0))$ is open and thus so is $f^{-1}(B_\epsilon(f(x_0)))$, which contains x_0 and therefore also (by openness) contains $B_\delta(x_0)$ for some $\delta > 0$.

Notice that conditions (b) and (c) do not require specific knowledge of the metric, but again only require knowing what an open subset is. Condition (b) is the one we will later use to define continuity in general topological spaces. It may be instructive to review why (b) and (c) are equivalent—especially because this is something that will turn out to be *false* in general for topological spaces, at least without some extra assumption.

Proof that (b) $\Leftrightarrow$ (c). To show that (b) $\Rightarrow$ (c), suppose $x_n \to x$ and $\mathcal{U} \subset Y$ is a neighborhood of $f(x)$. Then $\mathcal{U}$ contains an open set $\mathcal{V}$ containing $f(x)$, hence $f^{-1}(\mathcal{U})$ contains $f^{-1}(\mathcal{V})$ which contains x, and by condition (b), $f^{-1}(\mathcal{V})$ is also open, implying $f^{-1}(\mathcal{U})$ is a neighborhood of x. Convergence then implies that $x_n \in f^{-1}(\mathcal{U})$ and thus $f(x_n) \in \mathcal{U}$ for all n sufficiently large, which proves $f(x_n) \to f(x)$ since the neighborhood $\mathcal{U}$ was arbitrary.

For the other direction, we shall prove the contrapositive, i.e. we show that if (b) is false then so is (c). So assume there is an open subset $\mathcal{U} \subset Y$ such that $f^{-1}(\mathcal{U}) \subset X$ is not open. Being not open means that for some $x \in f^{-1}(\mathcal{U})$, no open ball about x is contained in $f^{-1}(\mathcal{U})$. As a consequence, for every $n \in \mathbb{N}$, we can find a point

$$x_n \in B_{1/n}(x) \quad \text{such that} \quad x_n \notin f^{-1}(\mathcal{U}),$$

meaning $f(x_n) \notin \mathcal{U}$. The sequence x_n then converges to x, since every neighborhood of x contains $B_{1/n}(x)$ for n sufficiently large, implying that x_n belongs to the given neighborhood for all large n. But $f(x_n)$ cannot converge to $f(x)$ since it never belongs to $\mathcal{U}$, which is a neighborhood of $f(x)$. □

I want to point out two things about the above proof. First, the proof that (b) $\Rightarrow$ (c) never mentioned the metric, it only talked about neighborhoods and open sets—as a consequence, that implication will remain true when we reconsider all these notions in general topological spaces. But the proof that (c) $\Rightarrow$ (b) did refer to the metric, because it used the precise definition of openness in terms of open balls. We will see that this implication does not actually hold in arbitrary topological spaces, though a mild modification of it does.

Definition 2.6. A map $f : X \to Y$ is a **homeomorphism** (*Homöomorphismus*) if it is continuous and bijective and its inverse $f^{-1} : Y \to X$ is also continuous.

Example 2.7. Consider $\mathbb{R}^n$ with the **standard Euclidean metric**

$$d_E(\mathbf{x}, \mathbf{y}) := |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}$$

for vectors $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ in $\mathbb{R}^n$. We claim that for any $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$, $(B_r(\mathbf{x}), d_E)$ is homeomorphic to $(\mathbb{R}^n, d_E)$. (It follows of course that all open balls in $\mathbb{R}^n$ are also homeomorphic to each other, though it is perhaps easier to prove the latter directly.) To construct a homeomorphism, choose any continuous, increasing, bijective function $f : [0, r) \to [0, \infty)$ and define $F : B_r(\mathbf{x}) \to \mathbb{R}^n$ by

$$F(\mathbf{x}) = \mathbf{x} \quad \text{and} \quad F(\mathbf{x} + \mathbf{y}) = \mathbf{x} + f(|\mathbf{y}|)\frac{\mathbf{y}}{|\mathbf{y}|} \text{ for all } \mathbf{y} \in B_r(0)\backslash\{0\} \subset \mathbb{R}^n.$$

It is easy to check that both F and F^{-1} are then continuous.

One conclusion to draw from the above example is that the notion of "boundedness," which is very important in analysis, is not going to make much sense in topology. Indeed, we would like to consider two spaces as "equivalent" whenever they are homeomorphic, so topologically it would be meaningless to call a space bounded if another space homeomorphic to it is not. What plays this role instead is the somewhat stricter notion of *compactness*. To write down the correct definition, we need to have the notion of an **open covering** (*offene Überdeckung*): assume I is any set (the so-called "index set") and $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is a collection of open subsets $\mathcal{U}_\alpha \subset X$ labeled by elements $\alpha \in I$. We call $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ an open covering/cover of a subset $A \subset X$ if

$$A \subset \bigcup_{\alpha \in I} \mathcal{U}_\alpha.$$

DEFINITION 2.8. A subset K in a metric space (X, d) is **compact** (*kompakt*) if either of the following equivalent conditions is satisfied:

(a) Every open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of K has a finite subcover (*eine endliche Teilüberdeckung*), i.e. there is a finite subset $\{\alpha_1, \ldots, \alpha_N\} \subset I$ such that

$$K \subset \bigcup_{i=1}^{N} \mathcal{U}_{\alpha_i}.$$

(b) Every sequence $x_n \in K$ has a convergent subsequence with limit in K.

We call (X, d) itself a **compact space** if X is a compact subset of itself.

Compactness is probably the least intuitive definition in this course so far, and at this stage we can only justify it by saying that it has stood the test of time: many beautiful and useful theorems have turned out to be true for compact spaces and *only* compact spaces. The first of these is the following, which explains why, unlike boundedness, compactness really is a topologically invariant notion, i.e. if X is compact, then so is every space that is homeomorphic to it.

THEOREM 2.9. *If $f : X \to Y$ is continuous and $K \subset X$ is compact, then so is $f(K) \subset Y$.*

PROOF. If $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open cover of $f(K)$, then the sets $f^{-1}(\mathcal{U}_\alpha)$ are all open in X and thus form an open cover of K, which is compact, so there is a finite subset $\{\alpha_1, \ldots, \alpha_N\} \subset I$ such that

$$K \subset \bigcup_{i=1}^{N} f^{-1}(\mathcal{U}_{\alpha_i}),$$

implying $f(K) \subset \bigcup_{i=1}^{N} \mathcal{U}_{\alpha_i}$, hence we have found a finite subcover of our given open cover of $f(K)$. □

One more remark about compactness: the equivalence of conditions (a) and (b) in Definition 2.8 is not so obvious, but is a fairly deep theorem called the *Bolzano-Weierstrass* theorem which you've probably seen proved in your analysis classes. We will prove an analogue of that theorem for topological spaces in Lecture 5, but it does not say that these two definitions are always equivalent—as with continuity, characterizing compactness via sequences becomes a slightly subtler issue in topological spaces, though the equivalence does hold for most of the spaces we actually care about.

Let's see some more examples now.

EXAMPLE 2.10. For any metric space (X, d) and an arbitrary subset $A \subset X$, (A, d) is also a metric space. So for instance, we can use the Euclidean metric d_E on $\mathbb{R}^{n+1}$ to define a metric on the subset

$$S^n = \left\{ \mathbf{x} \in \mathbb{R}^{n+1} \;\middle|\; |\mathbf{x}| = 1 \right\},$$

the n-dimensional sphere.

EXAMPLE 2.11. Any set X can be assigned the **discrete metric** (*diskrete Metrik*), defined by

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

This metric keeps every point at a measured distance away from every other point. So for instance, we can assign the discrete metric to $\mathbb{R}^n$ and compare it with the Euclidean metric d_E. We claim that the identity map on $\mathbb{R}^n$ defines a continuous map from $(\mathbb{R}^n, d_D)$ to $(\mathbb{R}^n, d_E)$, but it is not a homeomorphism, i.e. its inverse is not continuous. This follows immediately from the next exercise.

EXERCISE 2.12. Show that on any set X with the discrete metric d_D, every subset is open. In particular this includes the set $\{x\} \subset X$ for every $x \in X$. Conclude that a sequence x_n converges to x if and only if $x_n = x$ for all n sufficiently large, i.e. the sequence is "eventually constant". Then use this to prove the following statements:

(a) All maps from (X, d_D) to any other metric space are continuous.
(b) All continuous maps from $(\mathbb{R}^n, d_E)$ to (X, d_D) are constant.

EXAMPLE 2.13. Given two metric spaces (X, d_X) and (Y, d_Y), one can define a **product metric** on $X \times Y$ by

$$d_{X\times Y}((x,y),(x',y')) := \sqrt{d_X(x,x')^2 + d_Y(y,y')^2}.$$

This is the obvious generalization of the Euclidean metric, e.g. if X and Y are both $\mathbb{R}$ with its standard Euclidean metric, then $d_{X\times Y}$ becomes d_E on $\mathbb{R}^2$. But this is not the only reasonable choice of metric on $X \times Y$: for instance, one can also define a metric by

$$d'_{X\times Y}((x,y),(x',y')) := \max\left\{d_X(x,x'), d_Y(y,y')\right\}.$$

This metric is indeed different: for instance, if we again take X and Y to be the Euclidean $\mathbb{R}$, then an open ball with respect to $d'_{X\times Y}$ in $\mathbb{R}^2$ does not look circular, it looks rather like a square. On the other hand, this does not have a huge impact on the notion of open sets: it is not hard to show that the identity map from $(X \times Y, d_{X\times Y})$ to $(X \times Y, d'_{X\times Y})$ is always a homeomorphism.

DEFINITION 2.14. Two metrics d and d' on the same set X are called (topologically) **equivalent** if the identity map from (X, d) to (X, d') is a homeomorphism.

In light of the various ways we now have for defining what "continuous" means, equivalence of metrics can also be understood as follows:

- d and d' are equivalent if they both define the same notion of open subsets in X;
- d and d' are equivalent if they both define the same notion of convergence of sequences in X.

The characterization in terms of sequences is the subject of the next exercise.

EXERCISE 2.15. Suppose d_1 and d_2 are two metrics on the same set X. Show that the identity map defines a homeomorphism $(X, d_1) \to (X, d_2)$ if and only if the following condition is satisfied: for every sequence $x_n \in X$ and $x \in X$,

$$x_n \to x \text{ in } (X, d_1) \quad \Longleftrightarrow \quad x_n \to x \text{ in } (X, d_2).$$

EXAMPLE 2.16. In functional analysis, one often studies metric spaces whose elements are functions, and the exact choice of metric on such a space needs to be handled rather carefully. Consider for instance the set

$$X = C^0[-1,1] := \{\text{continuous functions } f : [-1,1] \to \mathbb{R}\}.$$

If we think of this as an infinite-dimensional vector space whose elements $f \in X$ are described by the (infinitely many) "coordinates" $f(t) \in \mathbb{R}$ for $t \in [-1, 1]$, then the natural generalization of the Euclidean metric to such a space is

$$d_2(f,g) := \sqrt{\int_{-1}^{1} |f(t) - g(t)|^2 \, dt}.$$

This is the metric corresponding to the so-called "L^2-norm" on the space of functions $[-1, 1] \to \mathbb{R}$. On the other hand, our alternative product metric discussed in Example 2.13 above generalizes to this space in the form

$$d_\infty(f,g) := \max_{t\in[-1,1]} |f(t) - g(t)|,$$

which is well defined since continuous functions on compact intervals always attain maxima. It is not hard to see that the identity map from (X, d_∞) to (X, d_2) is continuous, but is not a homeomorphism. Indeed, if $f_n \to f$ in (X, d_∞), then

$$d_2(f_n, f)^2 = \int_{-1}^{1} |f_n(t) - f(t)|^2\, dt \leqslant \int_{-1}^{1} \max_t |f_n(t) - f(t)|^2\, dt \leqslant 2 d_\infty(f_n, f)^2 \to 0,$$

proving that $f_n \to f$ also in (X, d_2). On the other hand, there exist sequences $f_n \in X$ such that $f_n \to 0$ with respect to d_2 but $d_\infty(f_n, 0) = 1$ for all n: just take a sequence of "bump" functions $f_n : [-1, 1] \to [0, 1]$ that all satisfy $f_n(0) = 1$ but vanish outside of progressively smaller neighborhoods of 0. These will satisfy $d_2(f_n, 0)^2 = \int_{-1}^{1} |f_n(t)|^2\, dt \to 0$, but $d_\infty(f_n, 0) = \max_t |f_n(t)| = 1$ for all n, preventing convergence to 0 with respect to d_∞.

Exercise 2.17. Suppose (X, d_X) is a metric space and $\sim$ is an equivalence relation on X, with the resulting set of equivalence classes denoted by $X/\sim$. For equivalence classes $[x], [y] \in X/\sim$, define

$$d([x], [y]) := \inf\left\{ d_X(x, y) \;\middle|\; x \in [x],\ y \in [y] \right\}. \tag{2.1}$$

(a) Show that d is a metric on $X/\sim$ if the following assumption is added: for every triple $[x], [y], [z] \in X/\sim$, there exist representatives $x \in [x]$, $y \in [y]$ and $z \in [z]$ such that

$$d_X(x, y) = d([x], [y]) \quad \text{and} \quad d_X(y, z) = d([y], [z]).$$

Comment: The hard part is proving the triangle inequality.

(b) Consider the real projective n-space

$$\mathbb{RP}^n := S^n/\sim,$$

where $S^n := \{\mathbf{x} \in \mathbb{R}^{n+1} \mid |\mathbf{x}| = 1\}$ and the equivalence relation identifies antipodal points, i.e. $\mathbf{x} \sim -\mathbf{x}$. If d_X is the metric on S^n induced by the standard Euclidean metric on $\mathbb{R}^{n+1}$, show that the extra assumption in part (a) is satisfied, so that (2.1) defines a metric on $\mathbb{RP}^n$.

(c) For the metric defined on $\mathbb{RP}^n$ in part (b), show that the natural quotient projection $\pi : S^n \to \mathbb{RP}^n$ sending each $\mathbf{x} \in S^n$ to its equivalence class $[\mathbf{x}] \in \mathbb{RP}^n$ is continuous, and a subset $\mathcal{U} \subset \mathbb{RP}^n$ is open if and only if $\pi^{-1}(\mathcal{U}) \subset S^n$ is open (with respect to the metric d_X).

(d) Here is a very different example of a quotient space. Define

$$X = (-1, 1)^2 \backslash \{(0, 0)\} \subset \mathbb{R}^2$$

with the metric d_X induced by the Euclidean metric on $\mathbb{R}^2$. Now fix the function $f : X \to \mathbb{R} : (x, y) \mapsto xy$ and define the relation $p_0 \sim p_1$ for $p_0, p_1 \in X$ to mean that there exists a continuous curve $\gamma : [0, 1] \to X$ with $\gamma(0) = p_0$ and $\gamma(1) = p_1$ such that $f \circ \gamma$ is constant. Show that for this equivalence relation, the extra assumption of part (a) is not satisfied, and the distance function defined in (2.1) does not satisfy the triangle inequality.

(e) Despite our failure to define $X/\sim$ as a metric space in part (d), it is natural to consider the following notion: define a subset $\mathcal{U} \subset X/\sim$ to be *open* if and only if $\pi^{-1}(\mathcal{U})$ is an open subset of (X, d_X), where $\pi : X \to X/\sim$ denotes the natural quotient projection. We can then define a sequence $[x_n] \in X/\sim$ to be *convergent* to an element $[x] \in X/\sim$ if for every open subset $\mathcal{U} \subset X/\sim$ containing $[x]$, $[x_n] \in \mathcal{U}$ for all n sufficiently large. Find a sequence $[x_n] \in X/\sim$ and two elements $[x], [y] \in X/\sim$ such that

$$[x_n] \to [x] \quad \text{and} \quad [x_n] \to [y], \quad \text{but} \quad [x] \neq [y].$$

This could not happen if we'd defined convergence on $X/\sim$ in terms of a metric. (Why not?)

Exercise 2.18.

(a) Show that for any metric space (X, d),

$$d'(x, y) := \min\{1, d(x, y)\}$$

defines another metric on X which is equivalent to d. In particular, this means that every metric is equivalent to one that is bounded.

(b) Suppose (X, d_X) and (Y, d_Y) are metric spaces satisfying

$$d_X(x, x') \leqslant 1 \text{ for all } x, x' \in X, \qquad d_Y(y, y') \leqslant 1 \text{ for all } y, y' \in Y.$$

Now let $Z = X \cup Y$, and for $z, z' \in Z$ define

$$d_Z(z, z') = \begin{cases} d_X(z, z') & \text{if } z, z' \in X, \\ d_Y(z, z') & \text{if } z, z' \in Y, \\ 2 & \text{if } (z, z') \text{ is in } X \times Y \text{ or } Y \times X. \end{cases}$$

Show that d_Z is a metric on Z with the following property: a subset $\mathcal{U} \subset Z$ is open in (Z, d_Z) if and only if it is the union of two (possibly empty) open subsets of (X, d_X) and (Y, d_Y). In particular, X and Y are each both open and closed subsets of Z. (Recall that subsets of metric spaces are closed if and only if their complements are open.)

(c) Suppose (Z, d) is a metric space containing two disjoint subsets $X, Y \subset Z$ that are each both open and closed. Show that there exists no continuous map $\gamma : [0, 1] \to Z$ with $\gamma(0) \in X$ and $\gamma(1) \in Y$.

(d) Show that if (X, d) is a metric space with the discrete metric, then for every point $x \in X$, the subset $\{x\} \subset X$ is both open and closed.

3. Topological spaces

We saw in the last lecture that most of the notions we want to consider in topology (continuous maps, homeomorphisms, convergence of sequences...) can be defined on metric spaces without specific reference to the metric, but using only our knowledge of which subsets are *open*. Moreover, one can define distinct but "equivalent" metrics on the same space for which the open sets match and therefore all these notions are the same. This suggests that we should view the notion of open sets as something more fundamental than a metric. The starting point of topology is to endow a set with the extra structure of a distinguished collection of subsets that we will call "open". The first question to answer is: what properties should we require this collection of subsets to have?

To motivate the axioms, let's revisit metric spaces for a moment and recall two important definitions. Both will also make sense in the context of topological spaces once we have fixed a definition for the latter.

Definition 3.1. Suppose X is a metric (or topological) space.

(a) The **interior** (*offener Kern* or *Inneres*) of a subset $A \subset X$ is the set

$$\mathring{A} = \left\{ x \in A \;\middle|\; \text{some neighborhood of } x \text{ in } X \text{ is contained in } A \right\}.$$

Points in this set are called **interior points** (*innere Punkte*) of A.

(b) The **closure** (*abgeschlossene Hülle* or *Abschluss*) of a subset $A \subset X$ is the set

$$\bar{A} = \left\{ x \in X \;\middle|\; \text{every neighborhood of } x \text{ in } X \text{ intersects } A \right\}.$$

Points in this set are called **cluster points** (*Berührpunkte*) of A.

The following exercise is easy, but it's worth thinking through why it is true.

Exercise 3.2. Show that for any subset $A \subset X$, the interior $\mathring{A}$ is the largest open subset of X that is contained in A, and the closure $\bar{A}$ is the smallest closed subset of X that contains A, i.e.

$$\mathring{A} = \bigcup_{\mathcal{U} \subset X \text{ open}, \ \mathcal{U} \subset A} \mathcal{U} \quad \text{and} \quad \bar{A} = \bigcap_{\mathcal{U} \subset X \text{ closed}, \ A \subset \mathcal{U}} \mathcal{U}.$$

I worded this exercise in a slightly sneaky way by calling the union of all the open sets inside A the "largest open subset of X that is contained in A": how do we actually know that this union of subsets is also open? This is the point: we know it because in a metric space, *arbitrary unions* of open subsets are also open. This follows almost immediately from the definitions in the previous lecture. It also implies (by taking complements) that arbitrary intersections of closed subsets are also closed, hence writing $\bar{A}$ as an intersection as in the exercise reveals that $\bar{A}$ is also a closed subset. These are properties you'd expect any reasonable notion of "open" or "closed" sets to have, so we will want to keep them.

What about intersections of open sets? Well, in metric spaces, arbitrary intersections of open sets need not be open, e.g. the intervals $(-1/n, 1/n) \subset \mathbb{R}$ are open for all $n \in \mathbb{N}$, but

$$\bigcap_{n \in \mathbb{N}} \left(-\frac{1}{n}, \frac{1}{n} \right) = \{0\}$$

is not an open subset of $\mathbb{R}$. Something slightly weaker is true, however: the intersection of any *two* open sets is open, and by an easy inductive argument, it follows that any *finite* intersection of open sets is open. Indeed, if $\mathcal{U}, \mathcal{V} \subset X$ are both open and $x \in \mathcal{U} \cap \mathcal{V}$, we know that $\mathcal{U}$ and $\mathcal{V}$ each contain balls about x for sufficiently small radii, so it suffices to take any radius small enough to fit inside both of them. (Why doesn't this necessarily work for an infinite intersection of open sets? Look at the example of the intervals $(-1/n, 1/n)$ above if you're not sure.) Taking complements, we also deduce from this discussion that arbitrary unions of closed subsets are not always closed, but *finite* unions are.

One last remark before we proceed: in any metric space X, the empty set $\varnothing$ and X itself are both open (and therefore also closed) subsets. With these observations as motivation, here is the definition on which everything else in this course will be based.

Definition 3.3. A **topology** (*Topologie*) on a set X is a collection[2] $\mathcal{T}$ of subsets of X satisfying the following axioms:

(i) $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$;
(ii) For every subcollection $I \subset \mathcal{T}$, $\bigcup_{\mathcal{U} \in I} \mathcal{U} \in \mathcal{T}$;
(iii) For every pair $\mathcal{U}_1, \mathcal{U}_2 \in \mathcal{T}$, $\mathcal{U}_1 \cap \mathcal{U}_2 \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is then called a **topological space** (*topologischer Raum*), and we call the sets $\mathcal{U} \in \mathcal{T}$ the **open** subsets (*offene Teilmengen*) in $(X, \mathcal{T})$.

We can now repeat several definitions from the previous lecture in our newly generalized context.

Definitions 3.4. Assume $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ are topological spaces.

(1) A subset $A \subset X$ is **closed** (*abgeschlossen*) if $X \backslash A \in \mathcal{T}_X$.

---

[2] I am calling $\mathcal{T}$ a "collection" instead of a "set" in an attempt to minimize the inevitable confusion caused by $\mathcal{T}$ being a set whose elements are also sets. Strictly speaking, there is nothing wrong with saying "$\mathcal{T}$ is a subset of 2^X satisfying the following axioms. . . ," where 2^X is the set-theoretician's fancy notation for the set consisting of all subsets of X. But if you found that sentence confusing, my recommendation is to call $\mathcal{T}$ a "collection" instead of a "set".

(2) A map $f : X \to Y$ is **continuous** (*stetig*) if for all $\mathcal{U} \in \mathcal{T}_Y$, $f^{-1}(\mathcal{U}) \in \mathcal{T}_X$. Note that if we prefer to describe the topology in terms of closed rather than open subsets, then it is equivalent to say that for all $\mathcal{U} \subset Y$ closed, $f^{-1}(\mathcal{U}) \subset X$ is also closed.

(3) A **neighborhood** (*Umgebung*) of a subset $A \subset X$ is any subset $\mathcal{U} \subset X$ such that $A \subset \mathcal{V} \subset \mathcal{U}$ for some $\mathcal{V} \in \mathcal{T}_X$.

(4) A sequence (*Folge*) $x_n \in X$ **converges to** (*konvergiert gegen*) $x \in X$ (written "$x_n \to x$") if for every neighborhood $\mathcal{U} \subset X$ of x, $x_n \in \mathcal{U}$ holds for all $n \in \mathbb{N}$ sufficiently large.

REMARK 3.5. One can equivalently define a topology $\mathcal{T}$ on a set X by specifying the *closed* sets $\mathcal{T}' := \{X \backslash \mathcal{U} \mid \mathcal{U} \in \mathcal{T}\}$. Then condition (ii) in Definition 3.3 is equivalent to

$$\bigcap_{A \in I} A \in \mathcal{T}' \quad \text{for all subcollections } I \subset \mathcal{T}',$$

and condition (iii) is equivalent to

$$A_1 \cup A_2 \in \mathcal{T}' \quad \text{for all } A_1, A_2 \in \mathcal{T}'.$$

For many topologies that one encounters in practice, it is not so easy to say what *all* the open sets look like, but much easier to describe a smaller subcollection that "generates" them.

DEFINITION 3.6. Suppose $(X, \mathcal{T})$ is a topological space and $\mathcal{B} \subset \mathcal{T}$ is a subcollection of the open sets.

- We call $\mathcal{B}$ a **base** or **basis** (*Basis*)[3] for $\mathcal{T}$ if every set $\mathcal{U} \in \mathcal{T}$ is a union of sets in $\mathcal{B}$, i.e.

$$\mathcal{U} = \bigcup_{\mathcal{V} \in I} \mathcal{V} \quad \text{for some subcollection } I \subset \mathcal{B}.$$

- We call $\mathcal{B}$ a **subbase** or **subbasis** (*Subbasis*) for $\mathcal{T}$ if every set $\mathcal{U} \in \mathcal{T}$ is a union of finite intersections of sets in $\mathcal{B}$, i.e.

$$\mathcal{U} = \bigcup_{\alpha \in I} \mathcal{U}_\alpha$$

for some collection of subsets $\mathcal{U}_\alpha \subset X$ indexed by a (possibly empty) set I, such that for each $\alpha \in I$,

$$\mathcal{U}_\alpha = \mathcal{U}_\alpha^1 \cap \ldots \cap \mathcal{U}_\alpha^{N_\alpha}$$

for some $N_\alpha \in \mathbb{N}$ and $\mathcal{U}_\alpha^1, \ldots, \mathcal{U}_\alpha^{N_\alpha} \in \mathcal{B}$.

Every base is obviously also a subbase, though we'll see in a moment that the converse is not true. You should take a moment to convince yourself that given any collection $\mathcal{B}$ of subsets of X that cover all of X (meaning $X = \bigcup_{\mathcal{U} \in \mathcal{B}} \mathcal{U}$), $\mathcal{B}$ is a subbase of a unique topology on X, namely the smallest topology that contains $\mathcal{B}$. It consists of all unions of finite intersections of sets from $\mathcal{B}$, and we say in this case that the topology $\mathcal{T}$ is **generated by** the collection $\mathcal{B}$.

EXAMPLE 3.7. The **standard topology** on $\mathbb{R}$ has the collection of all open intervals $\{(a, b) \subset \mathbb{R} \mid -\infty \leqslant a < b \leqslant \infty\}$ as a base. The smaller subcollection of half-infinite open intervals $\{(-\infty, a) \mid a \in \mathbb{R}\} \cup \{(a, \infty) \mid a \in \mathbb{R}\}$ is also a subbase, though not a base. (Why not?)

---

[3] Things got slightly confusing in Tuesday's lecture because when I stated the definition of a base, I neglected at first to require $\mathcal{B} \subset \mathcal{T}$, i.e. not only is every open set a union of sets from $\mathcal{B}$, but the sets in $\mathcal{B}$ are themselves also open, and as a result, *every* union of sets from $\mathcal{B}$ is also an open set. If one did not require the latter, then some stupid examples would be possible, e.g. the collection of one-point subsets would be a base for every topology. With the correct definition, however, $\mathcal{B}$ determines $\mathcal{T}$ uniquely, so taking $\mathcal{B}$ to consist of all one-point subsets automatically makes $\mathcal{T}$ the discrete topology.

EXAMPLE 3.8. If (X, d) is any metric (or pseudometric) space, the natural topology on X induced by the metric is defined via the base

$$\mathcal{B} = \left\{ B_r(x) \subset X \;\middle|\; x \in X,\ r > 0 \right\}.$$

Note that if d and d' are equivalent metrics as in Definition 2.14, then they induce the same topology on X: indeed, if the identity map $(X, d) \to (X, d')$ is a homeomorphism then it maps open sets to open sets. A topology that arises in this way from a metric is called **metrizable** (*metrisierbar*).

EXAMPLE 3.9. On any set X, the **discrete topology** is the collection $\mathcal{T}$ consisting of *all* subsets of X. Take a moment to convince yourself that this is a topology, and moreover, it is metrizable—it can be defined via the discrete metric, see Definition 2.11. (Can you think of another metric on X that defines the same topology?) As a base for $\mathcal{T}$, we can take $\mathcal{B} = \left\{ \{x\} \subset X \;\middle|\; x \in X \right\}$. Note that since all subsets are open, all subsets are also closed! Moreover:

- Every map $f : X \to \mathbb{R}$ is continuous.
- A map $f : \mathbb{R} \to X$ is continuous if and only if it is constant. Here is a quick proof: for every $x \in X$, $\{x\} \subset X$ is both open and closed, so continuity requires $f^{-1}(x) \subset \mathbb{R}$ also to be both open and closed, but the only subsets of $\mathbb{R}$ with this property are $\mathbb{R}$ itself and the empty set.
- A sequence $x_n \in X$ converges to $x \in X$ if and only if $x_n = x$ for all $n \in \mathbb{N}$ sufficiently large.

EXAMPLE 3.10. Also on any set X, one can define the **trivial** (also sometimes called the "indiscrete") topology $\mathcal{T} = \{\varnothing, X\}$. This topology has the distinguishing feature that every point $x \in X$ has only one neighborhood, namely the whole set. We then have:

- A map $f : X \to \mathbb{R}$ is continuous if and only if it is constant. Proof: Suppose f is continuous, $x_0 \in X$ and $f(x_0) = t \in \mathbb{R}$. Then for every $\epsilon > 0$, $f^{-1}(t - \epsilon, t + \epsilon)$ is an open subset of X containing x_0, so it is not $\varnothing$ and is therefore X. This proves

$$f(X) \subset \bigcap_{\epsilon > 0} (t - \epsilon, t + \epsilon) = \{t\}.$$

- All maps $f : \mathbb{R} \to X$ are continuous.
- $x_n \to x$ holds *always*, i.e. all sequences in X converge to all points! This proves that $(X, \mathcal{T})$ is not metrizable, as the limit of a convergent sequence in a metric space is always unique. (Prove it!)

EXAMPLE 3.11. The **cofinite** topology on a set X is defined such that a proper subset $A \subset X$ is closed if and only if it is finite. Take a moment to convince yourself that this really defines a topology—see Remark 3.5. (Note that X itself is automatically closed but does not need to be finite, since it is not a *proper* subset of itself.) The neighborhoods of a point $x \in X$ are then all of the form $X \backslash \{x_1, \ldots, x_N\}$ for arbitrary finite subsets $x_1, \ldots, x_N \in X$ that do not include x.

Suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on the same set X such that

$$\mathcal{T}_1 \subset \mathcal{T}_2,$$

meaning every open set in $(X, \mathcal{T}_1)$ is also an open set in $(X, \mathcal{T}_2)$. In this case we say that $\mathcal{T}_2$ is **stronger/finer/larger than** (*stärker/feiner als*) $\mathcal{T}_1$, and $\mathcal{T}_1$ is **weaker/coarser/smaller than** (*schwächer/gröber als*) $\mathcal{T}_2$. For example, since the open sets $\mathbb{R} \backslash \{x_1, \ldots, x_N\}$ for the cofinite topology on $\mathbb{R}$ are also open with respect to its standard topology, we can say that the standard topology of $\mathbb{R}$ is stronger than the cofinite topology. On any set, the discrete topology is the strongest, and the trivial topology is the weakest. In general, having a stronger topology means that fewer sequences converge, fewer maps into X from other spaces are continuous, but more functions defined on X are continuous. In various situations, it is common and natural to specify a topology on a set as being the "strongest" or "weakest" possible topology subject to the condition that some given collection of maps are all continuous. We will see some examples of this below.

There are several natural ways in which a given topology on one or more spaces can induce a topology on some related space.

Definition 3.12. $(X, \mathcal{T})$ determines on any subset $A \subset X$ the so-called **subspace topology** (*Unterraumtopologie*)

$$\mathcal{T}_A := \{\mathcal{U} \cap A \mid \mathcal{U} \in \mathcal{T}\}.$$

This is the weakest topology on A such that the natural inclusion $A \hookrightarrow X$ is a continuous map. (Prove it!)

Example 3.13. The standard topology on $\mathbb{R}^{n+1}$ is the one defined via the Euclidean metric. We then assign the subspace topology to the set of unit vectors $S^n \subset \mathbb{R}^{n+1}$, meaning a subset $\mathcal{V} \subset S^n$ will be considered open in S^n if and only if $\mathcal{V} = S^n \cap \mathcal{U}$ for some open subset $\mathcal{U} \subset \mathbb{R}^{n+1}$. As you might expect, this is the same as the topology induced by the metric on S^n defined by restricting the Euclidean metric, but for a given open set $\mathcal{V} \subset S^n$, it is not always so easy to see an open set $\mathcal{U} \subset \mathbb{R}^{n+1}$ such that $\mathcal{V} = \mathcal{U} \cap S^n$. Such a set can be constructed as follows: for each $\mathbf{x} \in \mathcal{V}$, choose $\epsilon_{\mathbf{x}} > 0$ such that every $\mathbf{y} \in S^n$ satisfying $|\mathbf{y} - \mathbf{x}| < \epsilon_{\mathbf{x}}$ is also in $\mathcal{V}$. Then the set

$$\mathcal{U} := \bigcup_{\mathbf{x} \in \mathcal{V}} \left\{\mathbf{y} \in \mathbb{R}^{n+1} \mid |\mathbf{y} - \mathbf{x}| < \epsilon_{\mathbf{x}}\right\}$$

is a union of open balls and is thus open in $\mathbb{R}^{n+1}$, and satisfies $\mathcal{U} \cap S^n = \mathcal{V}$.

Exercise 3.14. Convince yourself that for any metric space (X, d) and subset $A \subset X$, the natural metrizable topology on (A, d) is precisely the subspace topology with respect to the topology on X induced by d.

Definition 3.15. Given a collection of topological spaces $\{(X_\alpha, \mathcal{T}_\alpha)\}_{\alpha \in I}$ indexed by a set I such that $X_\alpha \cap X_\beta = \emptyset$ for all $\alpha \neq \beta$, the **disjoint union** (*disjunkte Vereinigung*) is the set $X := \bigcup_{\alpha \in I} X_\alpha$ with the topology

$$\mathcal{T} := \left\{ \bigcup_{\alpha \in I} \mathcal{U}_\alpha \;\middle|\; \mathcal{U}_\alpha \in \mathcal{T}_\alpha \text{ for all } \alpha \in I \right\}.$$

We typically denote the topological space $(X, \mathcal{T})$ defined in this way by

$$\coprod_{\alpha \in I} X_\alpha,$$

or for finite collections $I = \{1, \ldots, N\}$, $X_1 \amalg \ldots \amalg X_N$. The topology on this space is called the **disjoint union topology**.

Exercise 3.16. Show that the disjoint union topology $\mathcal{T}$ on $X = \coprod_\alpha X_\alpha$ is the strongest topology on this set such that for every $\alpha \in I$, the inclusion $X_\alpha \hookrightarrow X$ is continuous.

Remark 3.17. A key feature of the disjoint union topology is that for every individual $\alpha \in I$, the subset $X_\alpha \subset X$ is both open and closed. It follows that there is no continuous path $\gamma : [0,1] \to X$ with $\gamma(0) \in X_\alpha$ and $\gamma(1) \in X_\beta$ for $\alpha \neq \beta$, cf. Exercise 2.18(c).

Remark 3.18. It is also often useful to be able to discuss disjoint unions $\coprod_\alpha X_\alpha$ in which the sets X_α and X_β need not be disjoint for $\alpha \neq \beta$, e.g. a common situation is where all X_α are taken to be the same fixed set Y. In this case we still want to treat X_α and X_β as disjoint "copies" of the same subset when $\alpha \neq \beta$, so that no element in the union can belong to more than one of them. One way to do this is by redefining the set $X = \coprod_\alpha X_\alpha$ as

$$X := \left\{ (\alpha, x) \;\middle|\; \alpha \in I,\ x \in X_\alpha \right\},$$

so that the disjoint union topology now literally becomes the collection of all subsets in X of the form

$$\bigcup_{\alpha \in I} \{\alpha\} \times \mathcal{U}_\alpha$$

with $\mathcal{U}_\alpha \subset X_\alpha$ open for every α, and in analogy with Exercise 3.16, this is the strongest topology on X for which the injective maps $X_\alpha \to X : x \mapsto (\alpha, x)$ are continuous for all $\alpha \in I$. We will usually not bother with this cumbersome notation when examples arise: just remember that whenever X_1 and X_2 are two sets, disjoint or otherwise, the set $X_1 \amalg X_2$ is defined so that its subsets $X_1 \subset X_1 \amalg X_2$ and $X_2 \subset X_1 \amalg X_2$ are disjoint.

Exercise 3.19. Let $I = \mathbb{R}$ and define X_α for each $\alpha \in \mathbb{R}$ to be the same space consisting of only one element; for concreteness, say $X_\alpha := \{0\} \subset \mathbb{R}$. According to the definition described above, this sets up an obvious bijection

$$\coprod_{\alpha \in \mathbb{R}} \{0\} := \{(\alpha, 0) \in \mathbb{R} \times \{0\}\} \to \mathbb{R},$$
$$(\alpha, 0) \mapsto \alpha.$$

Show that this bijection is a homeomorphism if we assign the discrete topology to $\mathbb{R}$ on the right hand side.

4. Products, sequential continuity and nets

From now on, we'll adopt the following convention of terminology: if I say that X is a "**space**", then I mean X is a *topological* space unless I specifically say otherwise or the context clearly indicates that I mean something different (e.g. that X is a vector space). Similarly, if X and Y are spaces in the above sense and I refer to $f : X \to Y$ as a "**map**", then I typically mean that f is a *continuous* map unless the context indicates otherwise. We will sometimes have occasion to speak of maps $f : I \to X$ where X is a space but I is only a **set**, on which no topology has been specified: in this case no continuity is assumed since that notion is not well defined, but I will often try to be extra clear about it by calling f a "(not necessarily continuous) function" or something to that effect. I do not promise to be completely consistent about this, but hopefully my intended meaning will never be in doubt.

The previous lecture introduced two ways of inducing new topologies from old ones, namely on subspaces and on disjoint unions. It remains to discuss the natural topologies defined on products and quotients. We'll deal with the former in this lecture, and then use it to construct a surprising example illustrating the distinction between continuity and sequential continuity.

Definition 4.1. Given two spaces $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$, the **product topology** $\mathcal{T}$ on $X_1 \times X_2$ is generated by the base

$$\mathcal{B} := \left\{ \mathcal{U}_1 \times \mathcal{U}_2 \subset X_1 \times X_2 \;\middle|\; \mathcal{U}_1 \in \mathcal{T}_1,\ \mathcal{U}_2 \in \mathcal{T}_2 \right\}.$$

Notice that if $X_1 \times X_2$ is endowed with the product topology, then both of the projection maps

$$\pi_1 : X_1 \times X_2 \to X_1 : (x_1, x_2) \mapsto x_1$$
$$\pi_2 : X_1 \times X_2 \to X_2 : (x_1, x_2) \mapsto x_2$$

are continuous. Indeed, for any open set $\mathcal{U}_1 \subset X_1$, $\pi_1^{-1}(\mathcal{U}_1) = \mathcal{U}_1 \times X_2$ is the product of two open sets and is therefore open in $X_1 \times X_2$; similarly, $\pi_2^{-1}(\mathcal{U}_2) = X_1 \times \mathcal{U}_2$ is open if $\mathcal{U}_2 \subset X_2$ is open.

Notice moreover that the intersection of these two sets is $\mathcal{U}_1 \times \mathcal{U}_2$, so one can form all open sets in the product topology as unions of sets that are finite intersections of the form $\pi_1^{-1}(\mathcal{U}_1) \cap \pi_2^{-1}(\mathcal{U}_2)$. In other words, the subcollection

$$\{\pi_1^{-1}(\mathcal{U}) \mid \mathcal{U} \in \mathcal{T}_1\} \cup \{\pi_2^{-1}(\mathcal{U}) \mid \mathcal{U} \in \mathcal{T}_2\}$$

forms a subbase for the product topology $\mathcal{T}$. This makes $\mathcal{T}$ the weakest (i.e. smallest) topology for which the projection maps π_1 and π_2 are both continuous.

That last observation leads us to the natural generalization of this discussion to infinite products, but the outcome turns out to be slightly different from what you probably would have expected.

Suppose $\{(X_\alpha, \mathcal{T}_\alpha)\}_{\alpha \in I}$ is a collection of spaces, indexed by an arbitrary (possibly infinite) set I. Their product can be defined as the set

$$\prod_{\alpha \in I} X_\alpha := \left\{ \text{functions } f : I \to \bigcup_{\alpha \in I} X_\alpha : \alpha \mapsto x_\alpha \text{ such that } x_\alpha \in X_\alpha \text{ for all } \alpha \in I \right\}.$$

Note that since I in this discussion is only a set with no topology, there is no assumption of continuity for the functions $\alpha \mapsto x_\alpha$. Whether the set I is infinite or finite, we can denote elements of the product space by

$$\{x_\alpha\}_{\alpha \in I} \in \prod_{\alpha \in I} X_\alpha,$$

so we think of each of the individual elements $x_\alpha \in X_\alpha$ as "coordinates" on the product.

Definition 4.2. The **product topology** (*Produkttopologie*) on $\prod_{\alpha \in I} X_\alpha$ is the weakest topology such that all of the projection maps

$$\pi_\alpha : \prod_{\beta \in I} X_\beta \to X_\alpha : \{x_\beta\}_{\beta \in I} \mapsto x_\alpha$$

for $\alpha \in I$ are continuous.

In particular, the product topology must contain $\pi_\alpha^{-1}(\mathcal{U}_\alpha)$ for every $\alpha \in I$ and $\mathcal{U}_\alpha \in \mathcal{T}_\alpha$, and it is the smallest topology that contains them, which means the sets $\pi_\alpha^{-1}(\mathcal{U}_\alpha)$ form a subbase. It is important to spell out precisely what this means. We have

$$\pi_\alpha^{-1}(\mathcal{U}_\alpha) = \left\{ \{x_\beta\}_{\beta \in I} \in \prod_{\beta \in I} X_\beta \;\middle|\; x_\alpha \in \mathcal{U}_\alpha \right\},$$

so in each of these sets, only a single coordinate is constrained. It follows that in a finite intersection of sets of this form, only *finitely many* of the coordinates will be constrained, while the rest remain completely free. This implies:

Proposition 4.3. *A base for the product topology on $\prod_{\alpha \in I} X_\alpha$ is formed by the collection of all subsets of the form $\prod_{\alpha \in I} \mathcal{U}_\alpha$ where $\mathcal{U}_\alpha \subset X_\alpha$ is open for every $\alpha \in I$ and $\mathcal{U}_\alpha \neq X_\alpha$ is satisfied for at most finitely many $\alpha \in I$.* □

The last part of the above statement makes no difference when the product is finite, but for infinite products, it means that arbitrary subsets of the form $\prod_{\alpha \in I} \mathcal{U}_\alpha \subset \prod_{\alpha \in I} X_\alpha$ are not open just because $\mathcal{U}_\alpha \subset X_\alpha$ is open for every α. Dropping the "at most finitely many" condition would produce a much stronger topology with very different properties (see Exercise 4.6 below).

Exercise 4.4. Show that a sequence $\{x_\alpha^n\}_{\alpha \in I} \in \prod_{\alpha \in I} X_\alpha$ for $n \in \mathbb{N}$ converges as $n \to \infty$ to $\{x_\alpha\}_{\alpha \in I} \in \prod_{\alpha \in I} X_\alpha$ in the product topology if and only if for all $\alpha \in I$, the individual sequences x_α^n converge in X_α to x_α.

EXERCISE 4.5. Show that for any other space Y, a map $f : Y \to \prod_{\alpha \in I} X_\alpha$ is continuous if and only if $\pi_\alpha \circ f : Y \to X_\alpha$ is continuous for every $\alpha \in I$.

There is a special notation for the product set in the case where all the X_α are taken to be the same fixed space X: the product $\prod_{\alpha \in I} X$ has an obvious identification with the set of all (not necessarily continuous) functions $I \to X$, and we write

$$X^I := \prod_{\alpha \in I} X = \{\text{(not necessarily continuous) functions } f : I \to X\}.$$

For example we could now write $\mathbb{R}^n = \mathbb{R}^{\{1,\dots,n\}}$ if we preferred. The notation is motivated in part by the combinatorial observation that if X and I are both finite sets with a and b elements respectively, then X^I has a^b elements. The case $X = \{0,1\}$ is popular in abstract set theory since $\{0,1\}^I = \{f : I \to \{0,1\}\}$ has a straightforward interpretation as the set of all subsets of I, which is often abbreviated as $2^I := \{0,1\}^I$. But this example is not very interesting for topology since $\{0,1\}$ is not a very interesting topological space (no matter which topology you put on it—there are only four choices). When X is a more interesting space, the most important thing to understand about X^I comes from Exercise 4.4: a sequence of functions $f_n \in X^I$ converges to $f \in X^I$ if and only if it converges **pointwise**, i.e.

$$f_n(\alpha) \to f(\alpha) \quad \text{for every } \alpha \in I.$$

The product topology on X^I is therefore also sometimes called the **topology of pointwise convergence** (*punktweise Konvergenz*).

EXERCISE 4.6. Assume I is an infinite set and $\{(X_\alpha, \mathcal{T}_\alpha)\}_{\alpha \in I}$ is a collection of topological spaces. In addition to the usual product topology on $\prod_\alpha X_\alpha$, one can define the so-called *box topology*, which has a base of the form

$$\left\{ \prod_{\alpha \in I} \mathcal{U}_\alpha \;\middle|\; \mathcal{U}_\alpha \in \mathcal{T}_\alpha \text{ for all } \alpha \in I \right\}.$$

(a) Compared with the usual product topology, is the box topology stronger, weaker, or neither?

(b) What does it mean for a sequence in $\prod_\alpha X_\alpha$ to converge in the box topology? In particular, consider the case where all the X_α are a fixed space X and $\prod_\alpha X$ is identified with the space of all functions $X^I = \{f : I \to X\}$; what does it mean for a sequence of functions $f_n : I \to X$ to converge in the box topology to a function $f : I \to X$?

With examples like these at our disposal, we can now address the following important question in full generality:

QUESTION 4.7. *To what extent are the following conditions for maps $f : X \to Y$ between topological spaces equivalent?*

- $f^{-1}(\mathcal{U}) \subset X$ *is open for every open set* $\mathcal{U} \subset Y$*;*
- *For every convergent sequence $x_n \to x$ in X, $f(x_n) \to f(x)$ in Y.*

The first condition is ordinary continuity, while the second is called **sequential continuity** (*Folgenstetigkeit*). We proved in Lecture 2 that these two conditions are equivalent for maps between *metric spaces*, and if you look again at the proof that (b)⇒(c) in the discussion following Definition 2.5, you'll see that it still makes sense in arbitrary topological spaces, proving:

THEOREM 4.8. *For arbitrary topological spaces X and Y, all continuous maps $X \to Y$ are sequentially continuous.* □

The converse is trickier. Look again at the proof in Lecture 2 that (c)⇒(b) for Definition 2.5. That proof specifically referred to open balls about a point, so it is not so clear how to make sense of it in topological spaces where there is no metric. We can see however that the argument still works if we can remove all mention of open balls and replace it with the following lemma:

"Lemma" 4.9. *In any topological space X, a subset $A \subset X$ is not open if and only if there exists a point $x \in A$ and a sequence $x_n \in X \backslash A$ such that $x_n \to x$.*

I've put the word "lemma" in quotation marks here for a very good reason: as written, the statement is *false*, and so is the converse of Theorem 4.8! Sequential continuity does not always imply continuity. Here is a counterexample.

Example 4.10 (cf. [**Jän05**, §6.3]). Let $X = C^0([0,1],[-1,1]) \subset [-1,1]^{[0,1]}$, i.e. X is the set of all continuous functions $f : [0,1] \to [-1,1]$, and we assign to it the subspace topology as a subset of the space $[-1,1]^{[0,1]}$ of *all* functions $f : [0,1] \to [-1,1]$. In other words, X carries the topology of pointwise convergence. Next, define Y to be the same set, but with the topology induced by the L^2-metric

$$d_2(f,g) = \sqrt{\int_0^1 |f(t) - g(t)|^2 \, dt}.$$

Now consider the identity map from X to Y:

$$\Phi : X \to Y : f \mapsto f.$$

If $f_n \to f$ is a convergent sequence in X, then the functions converge pointwise, so $|f_n - f|^2$ converges pointwise to 0, and we claim that this implies $\int_0^1 |f_n(t) - f(t)|^2 \, dt \to 0$. This requires a fundamental result from measure theory, Lebesgue's *dominated convergence theorem* (see e.g. [**LL01**, §1.8] or [**Rud87**, Theorem 1.34]): it states that if g_n is a sequence of measurable functions that converge almost everywhere to g and all satisfy $|g_n| \leqslant G$ for some Lebesgue integrable function G, then $\int g_n$ converges to $\int g$. In the present case, the hypotheses are satisfied since the functions f_n take values in the bounded domain $[-1,1]$, which bounds $|f_n - f|$ uniformly below the constant (and thus integrable) function 2. We conclude that $d_2(f_n, f) \to 0$, hence Φ is sequentially continuous.

To show however that Φ is continuous, we would need to find for every $\epsilon > 0$ a neighborhood $\mathcal{U} \subset X$ of 0 such that $\Phi(\mathcal{U}) \subset B_\epsilon(0) \subset Y$. The trouble here is that neighborhoods in X (with the product topology) are somewhat peculiar objects: if $\mathcal{U}$ is one, then it contains some open set containing 0, which means it contains at least one of the sets $\prod_{\alpha \in [0,1]} \mathcal{U}_\alpha$ in our base for the product topology, where the $\mathcal{U}_\alpha$ are all open neighborhoods of 0 in $[-1,1]$ but there is at most a finite subset $I \subset [0,1]$ consisting of $\alpha \in [0,1]$ for which $\mathcal{U}_\alpha \neq [-1,1]$. Now choose a continuous function $f : [0,1] \to [0,1]$ that vanishes on the finite subset I but equals 1 on a "large" subset of $[0,1]\backslash I$. Depending how many points are in I, you may have to make this function oscillate very rapidly back and forth between 0 and 1, but since I is only finite, you can still do this such that the measure of the domain on which $f = 1$ is as close to 1 as you like, which makes $d_2(f,0)$ also only slightly less than 1. In particular, f belongs to the neighborhood $\mathcal{U}$ in X but not to $B_\epsilon(0) \subset Y$ if ϵ is sufficiently small.

We deduce from the above example that "Lemma" 4.9 is not always true, since it would imply that continuity and sequential continuity are equivalent. We are led to ask: what extra hypotheses could be added so that the lemma holds?

Definition 4.11. Given a point x in a space X, a **neighborhood base** (*Umgebungsbasis*) for x is a collection $\mathcal{B}$ of neighborhoods of x such that every neighborhood of x contains some $\mathcal{U} \in \mathcal{B}$.

Recall that a set I is **countable** (*abzählbar*) if it admits an injection into the natural numbers $\mathbb{N}$. This definition allows I to be either finite or infinite; if it is "countably infinite" then we can equivalently say that I admits a bijection with $\mathbb{N}$. This is also equivalent to saying that there exists a sequence $\{x_n \in I\}_{n\in\mathbb{N}}$ that includes every point of I. For example, it is easy to show that the set $\mathbb{Q}$ of rational numbers is countable, but Cantor's famous "diagonal" argument shows that $\mathbb{R}$ is not.

Definition 4.12 (the countability axioms). A space X is called **first countable** (*"X erfüllt das erste Abzählbarkeitsaxiom"*) if every point in x has a countable neighborhood base. We call X **second countable** (*"X erfüllt das zweite Abzählbarkeitsaxiom"*) if its topology has a countable base.

It is easy to see that every second countable space is also first countable: if X has a countable base $\mathcal{B}$, then for each $x \in X$, the collection of sets in $\mathcal{B}$ that contain x is a countable neighborhood base for x. The next example shows that the converse is false.

Example 4.13. If X has the discrete topology, then it is first countable because for each $x \in X$, one can form a neighborhood base out of the single open set $\{x\} \subset X$. But X is second countable if and only if X itself is a countable set (prove it!), so e.g. $\mathbb{R}$ with the discrete topology is first but not second countable.

Example 4.14. All metric spaces are first countable. Indeed, for every $x \in X$, the collection of open balls $B_{1/n}(x) \subset X$ for $n \in \mathbb{N}$ forms a countable neighborhood base. (Note that Example 4.13 is a special case of this, so not all metric spaces are second countable.)

We can now prove a corrected version of "Lemma" 4.9. Let us first make a useful general observation that follows directly from the axioms of a topology.

Lemma 4.15. *In any space X, a subset $A \subset X$ is open if and only if every point $x \in A$ has a neighborhood $\mathcal{V} \subset X$ that is contained in A.*

Proof. If the latter condition holds, then A is the union of open sets contained in such neighborhoods and is therefore open. Conversely, if A is open, then A itself can be taken as the desired neighborhood of every $x \in A$. ☐

Lemma 4.16. *In any first countable topological space X, a subset $A \subset X$ is not open if and only if there exists a point $x \in A$ and a sequence $x_n \in X \backslash A$ such that $x_n \to x$.*

Proof. If $A \subset X$ is open, then for every $x \in A$ and sequence $x_n \in X$ converging to x, we cannot have $x_n \in X\backslash A$ for all n since A is a neighborhood of x. This is true so far for *all* topological spaces, with or without the first countability axiom, but the latter will be needed in order to prove the converse. So, suppose now that $A \subset X$ is not open, which by Lemma 4.15, means there exists a point $x \in A$ such that no neighborhood $\mathcal{V} \subset X$ of x is contained in A. Fix a countable neighborhood base $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \ldots$ for x.

It will make our lives slightly easier if the neighborhood base is a nested sequence, meaning

$$X \supset \mathcal{U}_1 \supset \mathcal{U}_2 \supset \mathcal{U}_3 \supset \ldots \ni x,$$

and we claim that this can be assumed without loss of generality. Indeed, set $\mathcal{U}_1' := \mathcal{U}_1$, and if $\mathcal{U}_2$ is not contained in $\mathcal{U}_1'$, consider instead the set $\mathcal{U}_2 \cap \mathcal{U}_1'$, which is also a neighborhood of x and therefore (by the definition of a neighborhood base) contains $\mathcal{U}_n$ for some $n \in \mathbb{N}$. Since $\mathcal{U}_n$ is contained in $\mathcal{U}_1'$, we then set $\mathcal{U}_2' := \mathcal{U}_n$. Now continue this process by setting $\mathcal{U}_3' := \mathcal{U}_m$ such that $\mathcal{U}_m \subset \mathcal{U}_2' \cap \mathcal{U}_3$ and so forth. This algorithm produces a nested sequence $\mathcal{U}_1' \supset \mathcal{U}_2' \supset \mathcal{U}_3' \supset \ldots$ such that $\mathcal{U}_n' \subset \mathcal{U}_n$ for every n, hence the new neighborhoods also form a neighborhood base for x. Let us replace our original sequence with the nested sequence and continue to call it $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$.

With this new assumption in place, observe that since none of the neighborhoods $\mathcal{U}_n$ can be contained in A, there exists a sequence of points

$$x_n \in \mathcal{U}_n \quad \text{such that} \quad x_n \notin A.$$

This sequence converges to x since every neighborhood $\mathcal{V} \subset X$ of x contains one of the $\mathcal{U}_N$, implying that for all $n \geqslant N$,

$$x_n \in \mathcal{U}_n \subset \mathcal{U}_N \subset \mathcal{V}.$$

□

Combining this lemma with our proof in Lecture 2 that sequential continuity implies continuity in metric spaces yields:

COROLLARY 4.17. *For any spaces X and Y such that X is first countable, every sequentially continuous map $X \to Y$ is also continuous.* □

It is possible to generalize this result beyond first countable spaces, but it requires expanding our notion of what a "sequence" can be. If you think of a sequence in X as a map from the (ordered) set of natural numbers $\mathbb{N}$ to X, then one possible way to generalize is to consider more general partially ordered sets as domains. Recall that a binary relation $\prec$ defined on some subset of all pairs of elements in a set I is called a **partial order** (*Halbordnung* or *Teilordnung*) if it satisfies (i) $x \prec x$ for all x, (ii) $x \prec y$ and $y \prec x$ implies $x = y$, and (iii) $x \prec y$ and $y \prec z$ implies $x \prec z$. We write "$x \succ y$" as a synonym for "$y \prec x$", and the set I together with its partial order $\prec$ is called a **partially ordered set** (*partiell geordnete Menge*). One obvious example is $(\mathbb{N}, \leqslant)$, though unlike this example (which is *totally* ordered), it is not generally required in a partially ordered set $(I, \prec)$ that every pair of elements $x, y \in I$ satisfy either $x \prec y$ or $y \prec x$. We will see more exotic examples below.

DEFINITION 4.18. A **directed set** (*gerichtete Menge*) $(I, \prec)$ consists of a set I with a partial order $\prec$ such that for every pair $\alpha, \beta \in I$, there exists an element $\gamma \in I$ with $\gamma \succ \alpha$ and $\gamma \succ \beta$.

The natural numbers $(\mathbb{N}, \leqslant)$ clearly form a directed set, but in topology, one also encounters many interesting examples of directed sets that need not be totally ordered or countable.

EXAMPLE 4.19. If X is a space and $x \in X$, one can define a directed set $(I, \prec)$ where I is the set of all neighborhoods of x in X, and $\mathcal{U} \prec \mathcal{V}$ for $\mathcal{U}, \mathcal{V} \in I$ means $\mathcal{V} \subset \mathcal{U}$. This is a directed set because given any pair of neighborhoods $\mathcal{U}, \mathcal{V} \subset X$ of x, the intersection $\mathcal{U} \cap \mathcal{V}$ is also a neighborhood of x and thus defines an element of I with $\mathcal{U} \cap \mathcal{V} \subset \mathcal{U}$ and $\mathcal{U} \cap \mathcal{V} \subset \mathcal{V}$. Note that neither of $\mathcal{U}$ and $\mathcal{V}$ need be contained in the other, so they might not satisfy either $\mathcal{U} \prec \mathcal{V}$ or $\mathcal{V} \prec \mathcal{U}$.

DEFINITION 4.20. Given a space X, a **net** (*Netz*) $\{x_\alpha\}_{\alpha \in I}$ in X is a function $I \to X : \alpha \mapsto x_\alpha$, where $(I, \prec)$ is a directed set.

DEFINITION 4.21. We say that a net $\{x_\alpha\}_{\alpha \in I}$ in X **converges** to $x \in X$ if for every neighborhood $\mathcal{U} \subset X$ of x, there exists an element $\alpha_0 \in I$ such that $x_\alpha \in \mathcal{U}$ for every $\alpha \succ \alpha_0$.

Convergence of nets is also sometimes referred to in the literature as *Moore-Smith convergence*, see e.g. [**Kel75**]. Note that a net $\{x_\alpha\}_{\alpha \in I}$ whose underlying directed set is $(I, \prec) = (\mathbb{N}, \leqslant)$ is simply a sequence, and the above definition then reduces to the usual notion of convergence for a sequence. We can now prove the most general corrected version of "Lemma" 4.9.

LEMMA 4.22. *In any space X, a subset $A \subset X$ is not open if and only if there exists a point $x \in A$ and a net $\{x_\alpha\}_{\alpha \in I}$ in X that converges to x but satisfies $x_\alpha \notin A$ for every $\alpha \in I$.*

PROOF. If $A \subset X$ is open then it is a neighborhood of every $x \in A$, so the nonexistence of such a net is an immediate consequence of Definition 4.21. Conversely, if A is not open, then Lemma 4.15 provides a point $x \in A$ such that for every neighborhood $\mathcal{V} \subset X$ of x, there exists a point

$$x_{\mathcal{V}} \in \mathcal{V} \quad \text{such that} \quad x_{\mathcal{V}} \notin A.$$

Taking $(I, \prec)$ to be the directed set of all neighborhoods of x, ordered by inclusion as in Example 4.19, the collection of points $\{x_{\mathcal{V}}\}_{\mathcal{V} \in I}$ is now a net which converges to x since for every neighborhood $\mathcal{U} \subset X$ of x,

$$\mathcal{V} \succ \mathcal{U} \quad \Rightarrow \quad x_{\mathcal{V}} \in \mathcal{V} \subset \mathcal{U}.$$

□

Putting all this together leads to the following statement equating continuity with a generalized notion of sequential continuity. The proof is just a repeat of arguments we've already worked through, but we'll spell it out for the sake of completeness.

THEOREM 4.23. *For any spaces X and Y, a map $f : X \to Y$ is continuous if and only if for every net $\{x_\alpha\}_{\alpha \in I}$ in X converging to a point $x \in X$, the net $\{f(x_\alpha)\}_{\alpha \in I}$ in Y converges to $f(x)$.*

PROOF. Suppose f is continuous and $\{x_\alpha\}_{\alpha \in I}$ is a net in X converging to $x \in X$. Then for any neighborhood $\mathcal{U} \subset Y$ of $f(x)$, $f^{-1}(\mathcal{U}) \subset X$ is a neighborhood of x, hence there exists $\alpha_0 \in I$ such that $\alpha \succ \alpha_0$ implies $x_\alpha \in f^{-1}(\mathcal{U})$, or equivalently, $f(x_\alpha) \in \mathcal{U}$. This proves that $\{f(x_\alpha)\}_{\alpha \in I}$ converges in the sense of Definition 4.21 to $f(x)$.

To prove the converse, let us suppose that $f : X \to Y$ is not continuous, so there exists an open set $\mathcal{U} \subset Y$ for which $f^{-1}(\mathcal{U}) \subset X$ is not open. Then by Lemma 4.22, there exists a point $x \in f^{-1}(\mathcal{U})$ and a net $\{x_\alpha\}_{\alpha \in I}$ in X that converges to x but satisfies $x_\alpha \notin f^{-1}(\mathcal{U})$ for every $\alpha \in I$. Now $\{f(x_\alpha)\}_{\alpha \in I}$ is a net in Y that does not converge to $f(x)$, since $\mathcal{U}$ is an open neighborhood of $f(x)$ but $f(x_\alpha)$ is never in $\mathcal{U}$. □

Nets take a bit of getting used to in comparison with sequences. The following addendum to Example 4.10 may help in this regard, but it may also make you feel deeply unsettled.

EXAMPLE 4.24. For the identity map $\Phi : X \to Y$ in Example 4.10, one could extract from the above proof an example of a net $\{x_\alpha\}_{\alpha \in I}$ in X that converges to 0 without $\{\Phi(x_\alpha)\}_{\alpha \in I}$ converging to 0 in Y, but here is perhaps a slightly simpler example. Define I as the set of all finite subsets of $[0, 1]$, with the partial order $A \prec B$ for $A, B \subset [0, 1]$ defined to mean $A \subset B$. Note that $(I, \prec)$ is a directed set since for any two finite subsets $A, B \subset [0, 1]$, $A \cup B$ is also a finite subset and thus an element of I. Now choose for each $A \in I$ a continuous function

$$f_A : [0, 1] \to [0, 1]$$

such that $f_A|_A = 0$ but $\int_0^1 |f_A(t)|^2 \, dt > 1/4$. The net $\{\Phi(f_A)\}_{A \in I}$ in Y clearly does not converge to 0 since none of these functions belong to the ball $B_{1/2}(0)$ in Y. But $\{f_A\}_{A \in I}$ does converge to 0 in X: indeed, since X has the product topology, any neighborhood $\mathcal{U} \subset X$ of 0 contains some open neighborhood of 0 that is of the form $\prod_{\alpha \in [0,1]} \mathcal{U}_\alpha$ for open neighborhoods $\mathcal{U}_\alpha \subset [-1, 1]$ of 0 such that $\mathcal{U}_\alpha = [-1, 1]$ for all α outside of some finite subset $A_0 \subset [0, 1]$. It follows that for all $A \in I$ with $A \succ A_0 \in I$,

$$f_A(\alpha) = 0 \in \mathcal{U}_\alpha \text{ for all } \alpha \in A_0,$$

implying $f_A \in \mathcal{U}$.

5. Compactness

We saw in our discussion of metric spaces (Lecture 2) that boundedness is not a meaningful notion in topology, i.e. even if we have data such as a metric with which to define what a "bounded" set is, it may still be homeomorphic to sets that are not bounded. Instead, we consider *compact* sets, a notion that is topologically invariant. The main definition carries over from Lecture 2 with no change.

Definition 5.1. Given a space X and subset $A \subset X$, an **open cover/covering** (*offene Überdeckung*) of A is a collection of open subsets $\{\mathcal{U}_\alpha \subset X\}_{\alpha \in I}$ such that $A \subset \bigcup_{\alpha \in I} \mathcal{U}_\alpha$.

We will also occasionally use the notation

$$A \subset \bigcup_{\mathcal{U} \in \mathcal{O}} \mathcal{U}$$

to indicate an open covering of A, where $\mathcal{O}$ is a collection of open subsets of X, i.e. $\mathcal{O} \subset \mathcal{T}$, where $\mathcal{T}$ is the topology of X.

Definition 5.2. A subset $A \subset X$ is **compact** (*kompakt*) if every open cover of A has a finite subcover (*eine endliche Teilüberdeckung*), i.e. given an arbitrary open cover $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ of A, one can always find a finite subset $\{\alpha_1, \ldots, \alpha_N\} \subset I$ such that $A \subset \mathcal{U}_{\alpha_1} \cup \ldots \cup \mathcal{U}_{\alpha_N}$. We say that X itself is a **compact space** if X is a compact subset of itself.

Exercise 5.3. Show that a subset $A \subset X$ is compact if and only if A with the subspace topology is a compact space.

Example 5.4. For any space X with the discrete topology, a subset $A \subset X$ is compact if and only if A is finite. Indeed, the collection of subsets $\{\{x\} \subset X\}_{x \in A}$ forms an open covering of A in the discrete topology, and it has a finite subcovering if and only if A is finite, hence compactness implies finiteness. The converse follows from the next example.

Example 5.5. In any space X, every finite subset $A \subset X$ is compact. Indeed, for $A = \{a_1, \ldots, a_N\}$ with an open covering $\{\mathcal{U}_\alpha\}_{\alpha \in I}$, pick any $\alpha_i \in I$ with $a_i \in \mathcal{U}_{\alpha_i}$ for $i = 1, \ldots, N$, then the sets $\mathcal{U}_{\alpha_1}, \ldots, \mathcal{U}_{\alpha_N}$ form an open subcover.

Example 5.6. A subset $A \subset \mathbb{R}^n$ in Euclidean space with its standard topology is compact if and only if it is closed and bounded. This is known as the *Heine-Borel theorem*, and in one direction it is easy to prove; see Exercise 5.7 below. For the other direction, you have probably seen a proof in your analysis classes of the *Bolzano-Weierstrass theorem*, stating that if A is closed and bounded then every sequence in A has a convergent subsequence with limit in A; we say in this case that A is *sequentially compact*. We will prove in the following that compactness and sequential compactness are equivalent for second countable spaces, and every subset of $\mathbb{R}^n$ is second countable (see Exercise 5.9 below). A frequently occurring concrete example is the sphere

$$S^n \subset \mathbb{R}^{n+1},$$

which is a closed and bounded subset of $\mathbb{R}^{n+1}$ and is therefore compact.

Exercise 5.7. Show that in any metric space, compact subsets must be both closed and bounded.
Hint: For closedness, you may want to assume the theorem proved below that compact first countable spaces are also sequentially compact—recall that all metric spaces are first countable.

Remark 5.8. Note that the converse of Exercise 5.7 is generally false: being closed and bounded is not enough for compactness in arbitrary metric spaces. Here is an important class of examples from functional analysis: a vector space $\mathcal{H}$ with an inner product $\langle\ ,\ \rangle$ is called a **Hilbert**

space (*Hilbertraum*) if it is complete (meaning all Cauchy sequences converge) with respect to the metric $d(x, y) = \sqrt{\langle x - y, x - y\rangle}$. The closed unit ball $\bar{B}_1(0) = \{x \in \mathcal{H} \mid \langle x, x\rangle \leqslant 1\}$ is clearly both closed and bounded in $\mathcal{H}$, and it is compact if $\mathcal{H}$ is finite dimensional since, in this case, $\mathcal{H}$ is both linearly isomorphic and homeomorphic to $\mathbb{R}^n$ (or $\mathbb{C}^n$ in the complex case) with its standard inner product. But if $\mathcal{H}$ is infinite dimensional, then $\bar{B}_1(0)$ contains an infinite orthonormal set $e_1, e_2, e_3, \ldots$, i.e. satisfying

$$\langle e_i, e_i\rangle = 1 \text{ for all } i, \qquad \langle e_i, e_j\rangle = 0 \text{ if } i \neq j.$$

It then follows by a standard argument of Euclidean geometry that $d(e_i, e_j) = \sqrt{2}$ whenever $i \neq j$, so for any $r < \sqrt{2}/2$, no ball of radius r in $\mathcal{H}$ can contain more than one of these vectors. It follows that $\{B_r(x) \mid x \in \mathcal{H}\}$ is an open cover of $\bar{B}_1(0)$ that has no finite subcover. This way of characterizing the distinction between finite- and infinite-dimensional Hilbert spaces in terms of the compactness of the unit ball has useful applications, e.g. in the theory of elliptic PDEs. The latter has many quite deep applications in geometry and topology, for instance the index theory of Atiyah-Singer (see [**Boo77**, **BB85**]), gauge-theoretic invariants of smooth manifolds [**DK90**], and the theory of pseudoholomorphic curves in symplectic topology [**MS12**, **Wen18**].

Exercise 5.9. A space X is called **separable** (*separabel*) if it contains a countable subset $A \subset X$ that is also **dense** (*dicht*), meaning the closure[4] of A is X.

(a) Show that if X is a metric space and $A \subset X$ is a dense subset, then the collection of open balls $\{B_{1/n}(x) \subset X \mid n \in \mathbb{N},\ x \in A\}$ forms a base for the topology of X.
(b) Deduce that every separable and metrizable space is second countable.
(c) Show that $\mathbb{R}^n$ with its standard topology is separable.
(d) Show that if X is any second countable space, then every subset $A \subset X$ with the subspace topology is also second countable.

Example 5.10. A union of finitely many compact subsets in a space X is also compact. (This is an easy exercise.)

The next result implies that closed subsets in compact spaces are also compact.

Proposition 5.11. *For any compact subset $K \subset X$, if $A \subset X$ is closed and also is contained in K, then A is compact.*

Proof. Suppose $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open cover of A. Since A is closed, $X \backslash A$ is open, so that supplementing the collection $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ with $X \backslash A$ defines an open cover of X, and therefore also an open cover of K. Since K is compact, there is then a finite subset $\{\alpha_1, \ldots, \alpha_N\} \subset I$ such that

$$K \subset \mathcal{U}_{\alpha_1} \cup \ldots \cup \mathcal{U}_{\alpha_N} \cup (X \backslash A).$$

But $A \subset K$ is disjoint from $X \backslash A$, so this means $A \subset \mathcal{U}_{\alpha_1} \cup \ldots \cup \mathcal{U}_{\alpha_N}$, and we have found the desired finite subcover for A. ☐

The following theorem is just a repeat of Theorem 2.9, but in the more general context of topological rather than metric spaces. The proof carries over word for word.

Theorem 5.12. *If $f : X \to Y$ is continuous and $K \subset X$ is compact, then so is $f(K) \subset Y$.* ☐

Now would be a good moment to introduce the quotient topology, since it provides a large class of new examples of compact spaces.

---

[4] We gave the definition of the term *closure* in Lecture 3 (see Definition 3.1), originally in the context of metric spaces, but the same definition carries over to general topological spaces without change.

Definition 5.13. Suppose X is a space and $\sim$ is an equivalence relation on X, with the set of equivalence classes denoted by $X/\sim$. The **quotient topology** on $X/\sim$ is the strongest topology for which the natural projection map $\pi : X \to X/\sim$ sending each point $x \in X$ to its equivalence class $[x] \in X/\sim$ is continuous. Equivalently, a subset $\mathcal{U} \subset X/\sim$ is open in the quotient topology if and only if $\pi^{-1}(\mathcal{U})$ is an open subset of X.

I suggest you pause for a moment to make sure you understand why the two descriptions of the quotient topology in that definition are equivalent. Applying Theorem 5.12 to the continuous projection $\pi : X \to X/\sim$, we now have:

Corollary 5.14. *For any compact space X with an equivalence relation $\sim$, $X/\sim$ with the quotient topology is also compact.* ☐

Example 5.15. Since S^n is compact, so is $\mathbb{RP}^n = S^n/\{\mathbf{x} \sim -\mathbf{x}\}$ if we assign it the quotient topology. (Note that by Exercise 2.17(c), the quotient topology on $\mathbb{RP}^n$ is metrizable, and can be defined in terms of a natural metric induced on the quotient from the Euclidean metric restricted to S^n.)

Exercise 5.16. The space S^1, known as the **circle**, is normally defined as the unit circle in $\mathbb{R}^2$ and endowed with the subspace topology (induced by the Euclidean metric on $\mathbb{R}^2$). Show that the following spaces with their natural quotient topologies are both homeomorphic to S^1:

(a) $\mathbb{R}/\mathbb{Z}$, meaning the set of equivalence classes of real numbers where $x \sim y$ means $x - y \in \mathbb{Z}$.
(b) $[0, 1]/\sim$, where $0 \sim 1$.

For the next example, we introduce a convenient piece of standard notation. The quotient of a space X by a subset $A \subset X$ is defined as

$$X/A := X/\sim$$

with the quotient topology, where the equivalence relation is defined such that $x \sim y$ for every $x, y \in A$ and otherwise $x \sim x$ for all $x \in X$. In other words, X/A is the result of modifying X by "collapsing A to a point".

(c) Convince yourself that for every $n \in \mathbb{N}$, S^n is homeomorphic to $\mathbb{D}^n/S^{n-1}$, where

$$\mathbb{D}^n := \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x}| \leqslant 1\}.$$

Remark: Part (b) becomes a special case of part (c) if we replace $[0, 1]$ *by* $\mathbb{D}^1 = [-1, 1]$.

The remainder of this lecture will be concerned with the extent to which compactness is equivalent to the notion of **sequential compactness** (*Folgenkompaktheit*), defined as follows:

Definition 5.17. A subset $A \subset X$ is **sequentially compact** if every sequence in A has a subsequence that converges to a point in A.

As you might guess from our discussion of sequential continuity in the previous lecture, compactness and sequential compactness are not generally equivalent without some extra condition. But as with continuity, one obtains a result free of extra conditions by replacing sequences with *nets*.

Definition 5.18. Suppose $(I, \prec)$ is a directed set and $\{x_\alpha\}_{\alpha \in I}$ is a net in a space X. A point $x \in X$ is called a **cluster point** (*Häufungspunkt*) of $\{x_\alpha\}_{\alpha \in I}$ if for every neighborhood $\mathcal{U} \subset X$ of x and every $\alpha_0 \in I$, there exists $\alpha \succ \alpha_0$ such that $x_\alpha \in \mathcal{U}$.

Notice that the above definition is almost identical to that of *convergence* of $\{x_\alpha\}_{\alpha \in I}$ to x (see Definition 4.21), only the roles of "for every" and "there exist" have been reversed at the end. Informally, x being a cluster point does not require x_α to be arbitrarily close to x for *all* sufficiently large α, but only that one should be able to find *some* α arbitrarily large for which x_α is arbitrarily close. You should take a moment to think about what this definition means in the special case $(I, \prec) = (\mathbb{N}, \leqslant)$, where the net becomes a sequence, so the notion should be already familiar.

Definition 5.19. Given two directed sets $(I, \prec)$ and $(J, \prec)$, and nets $\{x_\alpha\}_{\alpha \in I}$ and $\{y_\beta\}_{\beta \in J}$ in a space X, we call $\{y_\beta\}_{\beta \in J}$ a **subnet** (*Teilnetz*) of $\{x_\alpha\}_{\alpha \in I}$ if $y_\beta = x_{\phi(\beta)}$ for all $\beta \in J$ and some function $\phi : J \to I$ with the property that for every $\alpha_0 \in I$, there exists $\beta_0 \in J$ for which $\beta \succ \beta_0$ implies $\phi(\beta) \succ \alpha_0$.

If $(I, \prec)$ and $(J, \prec)$ in the above definition are both $(\mathbb{N}, \leqslant)$ so that $\{x_\alpha\}_{\alpha \in I}$ and $\{y_\beta\}_{\beta \in I}$ become sequences x_n and y_k respectively, then y_k will be a subnet of x_n if it is of the form $y_k = x_{n_k}$ for some sequence $n_k \in \mathbb{N}$ satisfying $\lim_{k \to \infty} n_k = \infty$. This agrees with at least one of the standard definitions of the term **subsequence** (*Teilfolge*); a slightly stricter definition would require the sequence n_k to be monotone, but this difference is harmless. One should however be careful not to fall into the trap of thinking that a subnet of a sequence is always a subsequence—even if $(I, \prec) = (\mathbb{N}, \leqslant)$, Definition 5.19 allows much more general choices for the directed set $(J, \prec)$ and the function $\phi : J \to \mathbb{N}$ underlying a subnet of a sequence. In particular, the following lemma cannot be used to find convergent subsequences without imposing further conditions (cf. Lemma 5.22 below).

Lemma 5.20. *A net $\{x_\alpha\}_{\alpha \in I}$ in X has a cluster point at $x \in X$ if and only if it has a subnet convergent to x.*

Proof. Let us prove that a convergent subnet can always be derived from a cluster point x. Let $\mathcal{N}_x$ denote the set of all neighborhoods of x in X, and define $J = I \times \mathcal{N}_x$ with a partial order $\prec$ defined by

$$(\alpha, \mathcal{U}) \succ (\beta, \mathcal{V}) \quad \Leftrightarrow \quad \alpha \succ \beta \text{ and } \mathcal{U} \subset \mathcal{V}.$$

This makes $(J, \prec)$ a directed set since $(I, \prec)$ is already a directed set and the intersection of two neighborhoods is a neighborhood contained in both. Now since x is a cluster point of the net $\{x_\alpha\}_{\alpha \in I}$, there exists a function $\phi : J \to I$ such that for all $(\beta, \mathcal{U}) \in J$, $\phi(\beta, \mathcal{U}) =: \alpha$ satisfies $\alpha \succ \beta$ and $x_\alpha \in \mathcal{U}$. It is then straightforward to check that $\{x_{\phi(\beta,\mathcal{U})}\}_{(\beta,\mathcal{U}) \in J}$ is a subnet convergent to x.

The converse is easier, so I will leave it as an exercise. □

Here is the most general result relating compactness to nets.

Theorem 5.21. *A space X is compact if and only if every net in X has a convergent subnet.*

Proof. We prove first that if X is compact, then every net $\{x_\alpha\}_{\alpha \in I}$ has a cluster point (and therefore by Lemma 5.20 a convergent subnet). Arguing by contradiction, suppose no $x \in X$ is a cluster point of $\{x_\alpha\}_{\alpha \in I}$. Then one can associate to every $x \in X$ a neighborhood $\mathcal{U}_x$ and an element $\alpha_x \in I$ such that for every $\alpha \succ \alpha_x$, $x_\alpha \notin \mathcal{U}_x$. Without loss of generality let us suppose the neighborhoods $\mathcal{U}_x$ are all open. Then the collection of sets $\{\mathcal{U}_x\}_{x \in X}$ forms an open cover of X, and therefore has a finite subcover since X is compact. This means there is a finite set of points $x_1, \ldots, x_N \in X$ such that $X = \mathcal{U}_{x_1} \cup \ldots \cup \mathcal{U}_{x_N}$. Now since $(I, \prec)$ is a directed set, we can find an element $\beta \in I$ satisfying

$$\beta \succ \alpha_{x_i} \text{ for all } i = 1, \ldots, N,$$

hence $x_\beta \notin \mathcal{U}_{x_i}$ for every $i = 1, \ldots, N$. But the latter sets cover X, so this is impossible, and we have found a contradiction.

For the converse, we shall prove that if X is not compact then there exists a net with no cluster point. Being noncompact means one can find a collection $\mathcal{O}$ of open subsets such that $X = \bigcup_{\mathcal{U} \in \mathcal{O}} \mathcal{U}$ but no finite subcollection of them has union equal to X. Define I to be the set of all finite subcollections of the sets in $\mathcal{O}$, so by assumption, one can associate to every $\mathcal{A} \in I$ a point $x_{\mathcal{A}} \in X$ satisfying

$$x_{\mathcal{A}} \notin \bigcup_{\mathcal{U} \in \mathcal{A}} \mathcal{U}. \tag{5.1}$$

Define a partial order $\prec$ on I by

$$\mathcal{A} \prec \mathcal{B} \quad \Leftrightarrow \quad \mathcal{A} \subset \mathcal{B},$$

and notice that $(I, \prec)$ is now a directed set since the union of any two finite subcollections is another finite subcollection that contains both. This makes $\{x_{\mathcal{A}}\}_{\mathcal{A} \in I}$ a net in X, and we claim that it has no cluster point. Indeed, if $x \in X$ is a cluster point of $\{x_{\mathcal{A}}\}_{\mathcal{A} \in I}$, then since the sets in $\mathcal{O}$ cover X, there is a set $\mathcal{V} \in \mathcal{O}$ that is a neighborhood of x, and it follows that there must exist some $\mathcal{A} \succ \{\mathcal{V}\}$ in I for which

$$x_{\mathcal{A}} \in \mathcal{V} \subset \bigcup_{\mathcal{U} \in \mathcal{A}} \mathcal{U}.$$

This contradicts (5.1) and thus proves the claim that there is no cluster point. □

The next step is to impose countability axioms so that Theorem 5.21 gives us corollaries about sequential compactness.

Lemma 5.22. *If $x_n \in X$ is a sequence with a cluster point at $x \in X$ and x has a countable neighborhood base, then x_n has a subsequence converging to x.*

Proof. As in the proof of Lemma 4.16, we can assume without loss of generality that our countable neighborhood base has the form of a nested sequence of neighborhoods

$$X \supset \mathcal{U}_1 \supset \mathcal{U}_2 \supset \ldots \ni x.$$

Since x is a cluster point, we can choose $k_1 \in \mathbb{N}$ so that $x_{k_1} \in \mathcal{U}_1$, and then inductively for each $n \in \mathbb{N}$, choose $k_n \in \mathbb{N}$ such that $x_{k_n} \in \mathcal{U}_n$ and $k_n > k_{n-1}$. Then x_{k_n} is a subsequence of x_n and it converges to x, since for all neighborhoods $\mathcal{V} \subset X$ of x, we have $\mathcal{V} \supset \mathcal{U}_N$ for some $N \in \mathbb{N}$, implying

$$n \geqslant N \quad \Rightarrow \quad x_{k_n} \in \mathcal{U}_n \subset \mathcal{U}_N \subset \mathcal{V}.$$

□

Corollary 5.23. *If X is compact and first countable, then it is also sequentially compact.* □

Example 5.24. Though it is not so easy to see this, the space $[0,1]^{\mathbb{R}}$ of (not necessarily continuous) functions $\mathbb{R} \to [0,1]$ with the topology of pointwise convergence is compact, but not sequentially compact. Compactness follows directly from a deep result known as Tychonoff's theorem, which we will discuss in the next lecture. For the construction of a sequence in $[0,1]^{\mathbb{R}}$ with no convergent subsequence, see Exercise 6.5.

To prove compactness from sequential compactness, it turns out that we will need to invoke the second countability axiom. In practice, almost all of the spaces that topologists spend their time thinking about are second countable, resulting from the fact that most of them are separable and metrizable (see Exercise 5.9). One useful property shared by all second countable (but not necessarily compact) spaces is the following.

Lemma 5.25. *If X is second countable, then every open cover of X has a countable subcover.*

Proof. Assume $\{\mathcal{U}_\alpha\}_{\alpha \in I}$ is an open cover of X and $\mathcal{B}$ is a countable base. Then each $\mathcal{U}_\alpha$ is a union of sets in $\mathcal{B}$, and the collection of all sets in $\mathcal{B}$ that are contained in some $\mathcal{U}_\alpha$ is a countable subcollection $\mathcal{B}' \subset \mathcal{B}$ that also covers X. Let us denote $\mathcal{B}' = \{\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \ldots\}$. We can now choose for each $\mathcal{V}_n \in \mathcal{B}'$ an element $\alpha_n \in I$ such that $\mathcal{V}_n \subset \mathcal{U}_{\alpha_n}$, and $\{\mathcal{U}_{\alpha_n}\}_{n \in \mathbb{N}}$ is then a countable subcover of $\{\mathcal{U}_\alpha\}_{\alpha \in I}$. □

If you now take the second half of the proof of Theorem 5.21 and redo it with the focus on sequences instead of nets, and with Lemma 5.25 in mind, the result is the following.

THEOREM 5.26. *If X is second countable and sequentially compact, then it is compact.*

PROOF. We need to show that every open cover of X has a finite subcover. Since X is second countable, we can first use Lemma 5.25 to reduce the given open cover to a *countable* subcover $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \ldots \subset X$. Now arguing by contradiction, suppose that X is sequentially compact but the sets $\mathcal{U}_1, \ldots, \mathcal{U}_n$ do not cover X for any $n \in \mathbb{N}$, hence there exists a sequence $x_n \in X$ such that

$$x_n \notin \mathcal{U}_1 \cup \ldots \cup \mathcal{U}_n \tag{5.2}$$

for every $n \in \mathbb{N}$. Some subsequence x_{k_n} then converges to a point $x \in X$, which necessarily lies in $\mathcal{U}_N$ for some $N \in \mathbb{N}$. It follows that x_{k_n} also lies in $\mathcal{U}_N$ for all n sufficiently large, but this contradicts (5.2) as soon as $k_n \geqslant N$. □

EXERCISE 5.27. Consider the space

$$X = \left\{ f \in [0,1]^{\mathbb{R}} \;\middle|\; f(x) \neq 0 \text{ for at most countably many points } x \in \mathbb{R} \right\},$$

with the subspace topology that it inherits from $[0,1]^{\mathbb{R}}$.

(a) Show that X is sequentially compact.
Hint: For any sequence $f_n \in X$, the set $\bigcup_{n \in \mathbb{N}} \{x \in \mathbb{R} \mid f_n(x) \neq 0\}$ is also countable.

(b) For each $x \in \mathbb{R}$, define $\mathcal{U}_x = \{f \in X \mid -1 < f(x) < 1\}$. Show that the collection $\{U_x \subset X \mid x \in \mathbb{R}\}$ forms an open cover of X that has no finite subcover, hence X is not compact.

Corollary 5.23 and Theorem 5.26 combine to give the following result that is easy to remember:

COROLLARY 5.28. *A second countable space is compact if and only if it is sequentially compact.* □

A loose end: We know from Exercise 5.9 that every separable metric space is second countable, thus Corollary 5.28 implies the equivalence of compactness and sequential compactness for separable metric spaces, which includes most of the metric spaces that one uses in practice. However, more than this was claimed in Lecture 2: the equivalence should hold in *all* metric spaces, and this does not quite follow from what we've proved here. The missing ingredient needed is the notion of *total boundedness*: one can show that every sequentially compact set A in a metric space X is **totally bounded** (*total beschränkt*), meaning that for every $\epsilon > 0$, A is contained in the union of finitely many balls of radius ϵ. Taking $\epsilon = 1/n$ for $n \in \mathbb{N}$ then provides a countable collection of open balls covering A, which can serve as a substitute for the countable subcover we used in the proof of Theorem 5.26. We will not go further into the details here, since this is a topology and not an analysis course, and we will not need the result going forward.

6. Tychonoff's theorem and the separation axioms

Topic 1: Products of compact spaces. Here is a result that may sound less surprising at first than it actually is.

THEOREM 6.1 (Tychonoff's theorem). *For any collection of compact spaces $\{X_\alpha\}_{\alpha \in I}$, the product $\prod_{\alpha \in I} X_\alpha$ is compact.*

NONMATHEMATICAL REMARK. Thinking like an Anglophone may lead you to false assumptions about the pronunciation of the name Tychonoff, e.g. I was mispronouncing it for years until I finally looked up the name on Wikipedia in the context of teaching this course. The original Russian spelling is Тихонов, which would normally get transliterated into English as Tikhonov. The reason he instead became known outside of Russia as Tychonoff is that his papers were published in German, hence different phonetic conventions.

When I is a finite set, Theorem 6.1 says something not at all surprising, and the proof is straightforward, so let's start with that.

PROOF OF THEOREM 6.1 FOR FINITE PRODUCTS. By induction, it will suffice to prove that if X and Y are both compact spaces then so is $X \times Y$. We will do so by showing that every net in $X \times Y$ has a convergent subnet. Recall that a net $\{(x_\alpha, y_\alpha)\}_{\alpha\in I}$ in $X \times Y$ converges to $(x, y) \in X \times Y$ if and only if the nets $\{x_\alpha\}_{\alpha\in I}$ in X and $\{y_\alpha\}_{\alpha\in I}$ in Y converge to x and y respectively. (The corresponding fact about sequences was proved in Exercise 4.4—the proof for nets is the same.) Now, since X is compact, $\{x_\alpha\}_{\alpha\in I}$ has a subnet $\{x_{\phi(\beta)}\}_{\beta\in J}$ convergent to some point $x \in X$, where J is some other directed set with a suitable function $\phi : J \to I$. Compactness of Y implies in turn that $\{y_{\phi(\beta)}\}_{\beta\in J}$ has a subnet $\{y_{\phi(\psi(\gamma))}\}_{\gamma\in K}$ convergent to some point $y \in Y$. We therefore obtain a subnet

$$\{(x_{\phi\circ\psi(\gamma)}, y_{\phi\circ\psi(\gamma)})\}_{\gamma\in K}$$

of the original net $\{(x_\alpha, y_\alpha)\}_{\alpha\in I}$ that converges in $X \times Y$ to (x, y). □

The much less obvious aspect of Theorem 6.1 is that it is also true for infinite products, even those for which the index set I is *uncountably* infinite. So it follows for instance that the space

$$[0,1]^{\mathbb{R}} = \{\text{not necessarily continuous functions } f : \mathbb{R} \to [0,1]\} = \prod_{\alpha\in\mathbb{R}} [0,1]$$

with the topology of pointwise convergence is compact, as an immediate consequence of the fact that $[0, 1]$ is compact. Of course, this does not mean that every sequence of functions $f_n : \mathbb{R} \to [0, 1]$ has a pointwise convergent subsequence! That would be truly surprising, but it is false (see Exercise 6.5); it turns out that $[0, 1]^{\mathbb{R}}$ is not a first countable space, so it is allowed to be compact without being sequentially compact.

For a slightly different example, $[-1, 1]^{\mathbb{N}}$ is compact. We can identify this space with the set of all sequences in $[-1, 1]$, again with the topology of pointwise convergence, i.e. a sequence of sequences $\{x^n_k\}_{k\in\mathbb{N}} \in [-1, 1]^{\mathbb{N}}$ converges as $n \to \infty$ to a sequence $\{x_k\}_{k\in\mathbb{N}}$ if $\lim_{n\to\infty} x^n_k = x_k$ for every $k \in \mathbb{N}$. Now observe that $[-1, 1]^{\mathbb{N}}$ also contains the unit ball in the infinite-dimensional Hilbert space

$$\ell^2[-1,1] := \left\{ \{x_k \in \mathbb{R}\}_{k\in\mathbb{N}} \;\middle|\; \sum_{k=1}^{\infty} |x_k|^2 < \infty \right\}$$

with metric defined by

$$d(\{x_k\}, \{y_k\})^2 = \sum_{k=1}^{\infty} |x_k - y_k|^2.$$

The unit ball in $\ell^2[-1, 1]$ is clearly noncompact since it contains the sequence of sequenes

$$(1, 0, 0, \ldots), (0, 1, 0, \ldots), (0, 0, 1, 0, \ldots), \ldots,$$

which converges pointwise to 0 but stays at a constant distance away from 0 with respect to the metric, so it can have no convergent subsequence in the topology of $\ell^2[-1, 1]$. It may seem surprising in this case that the *larger* set $[-1, 1]^{\mathbb{N}}$ is compact, but the reason is that $[-1, 1]^{\mathbb{N}}$ has a much weaker topology than $\ell^2[-1, 1]$: since it is easier to converge pointwise than it is to converge in the ℓ^2-norm, $[-1, 1]^{\mathbb{N}}$ has more sequences with convergent subsequences (or subnets, as the case may be).

REMARK 6.2. One conclusion you should draw from the above discussion is that Tychonoff's theorem depends crucially on the way we defined the product topology on $\prod_{\alpha \in I} X_\alpha$, i.e. it is a result about the topology of pointwise convergence. The result becomes false, for instance, if we replace the usual product topology by the "box" topology from Exercise 4.6. For a concrete example, consider the set $[-1,1]^{\mathbb{N}}$ with the box topology, meaning sets of the form

$$\left\{ f \in [-1,1]^{\mathbb{N}} \;\middle|\; f(k) \in \mathcal{U}_k \text{ for all } k \in \mathbb{N} \right\}$$

for arbitrary collections of open subsets $\{\mathcal{U}_k \subset [-1,1]\}_{k \in \mathbb{N}}$ are open. Then the sequence of constant functions $f_n(k) := 1/n$ converges pointwise to 0, but we claim that it has no cluster point in the box topology. Indeed, the box topology contains the product topology, so if any subnet of f_n converges in the box topology, then it must also converge in the product topology and hence pointwise, meaning the only limit it could possibly converge to is 0, and 0 is therefore the only possible cluster point. But in the box topology,

$$\mathcal{U} := \left\{ f \in [-1,1]^{\mathbb{N}} \;\middle|\; f(k) \in (-1/k, 1/k) \text{ for all } k \in \mathbb{N} \right\}$$

is an open neighborhood of 0 satisfying $f_n \notin \mathcal{U}$ for all $n \in \mathbb{N}$, so 0 is not a cluster point of this sequence.

Let's go ahead and prove another special case of Tychonoff's theorem. The next proof is still relatively straightforward, and it applies for instance to $[-1,1]^{\mathbb{N}}$. Part of the idea is to make our lives easier by dealing with sequences instead of nets, which is made possible by the following simple observation:

LEMMA 6.3. *If $X_1, X_2, X_3, \ldots$ is a countably infinite sequence of spaces that are all second countable, then $\prod_{i=1}^{\infty} X_i$ is also second countable.*

PROOF. Fix for each $i = 1, 2, 3, \ldots$ a countable base $\mathcal{B}_i$ for the topology of X_i. Then for each $n \in \mathbb{N}$, the collection of sets

$$\mathcal{O}_n := \left\{ \mathcal{U}_1 \times \ldots \times \mathcal{U}_n \times X_{n+1} \times X_{n+2} \times \ldots \subset \prod_{i=1}^{\infty} X_i \;\middle|\; \mathcal{U}_i \in \mathcal{B}_i \text{ for each } i = 1, \ldots, n \right\}$$

is countable since $\mathcal{B}_1 \times \ldots \times \mathcal{B}_n$ is countable. Then the countable union of countable sets $\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{O}_3 \cup \ldots$ is a base for $\prod_{i=1}^{\infty} X_i$, and it is countable. □

PROOF OF THEOREM 6.1, SECOND COUNTABLE CASE. Assume the set I is countable and the spaces X_α are all second countable for $\alpha \in I$. In light of Lemma 6.3 and Theorem 5.26, it will now suffice to prove that for any sequence $X_1, X_2, X_3, \ldots$ of second countable spaces, $\prod_{i=1}^{\infty} X_i$ is sequentially compact. The idea is to combine the argument above for the case of finite products with Cantor's diagonal method. In order to avoid too many indices, let us denote elements $f \in \prod_{i=1}^{\infty} X_i$ as functions $f : \mathbb{N} \to \bigcup_{i=1}^{\infty} X_i$ that satisfy $f(i) \in X_i$ for each $i \in \mathbb{N}$. Now given a sequence $f_n \in \prod_{i=1}^{\infty} X_i$, the compactness of X_1 guarantees that there is a subsequence f_n^1 of f_n for which the sequence $f_n^1(1)$ in X_1 converges. Continuing inductively, we can construct a sequence of sequences $f_n^k \in \prod_{i=1}^{\infty} X_i$ for $k, n \in \mathbb{N}$ such that for every $k \geqslant 2$, $\{f_n^k\}_{n=1}^{\infty}$ is a subsequence of $\{f_n^{k-1}\}_{n=1}^{\infty}$ and the sequence $f_n^k(k)$ in X_k converges as $n \to \infty$. It follows that for every fixed $k \in \mathbb{N}$, the sequence $\{f_n^n(k)\}_{n=1}^{\infty}$ in X_k converges, thus $\{f_n^n\}_{n=1}^{\infty}$ is a convergent subsequence of the original sequence f_n in $\prod_{i=1}^{\infty} X_i$. □

The ideas in the special cases we've treated so far can be applied toward a general proof of Tychonoff's theorem, but the general case requires one major ingredient that wasn't needed so far: the axiom of choice. This makes e.g. the compactness of $[-1,1]^{[0,1]}$ somewhat harder to grasp intuitively, as invoking the axiom of choice means that the existence of a cluster point for every sequence in $[-1,1]^{[0,1]}$ is guaranteed, but there is nothing even slightly resembling an algorithm for finding one. It is known in fact that this is not just a feature of any particular method of proving the theorem—by a result due to Kelley [**Kel50**], if one assumes that the usual axioms of set theory (not including choice) hold and that Tychonoff's theorem also holds, then the axiom of choice follows, thus the two are actually equivalent.

Speaking only for myself, I had a Ph.D. in mathematics already for several years before I ever started to find the axiom of choice remotely worrying, so if you've never worried about it before, I don't encourage you to start worrying now. As far as this course is concerned, we actually could have skipped the general case of Tychonoff's theorem with no significant loss of continuity—I am including it here mainly for the sake of cultural education, and because the proof itself is interesting.

The proof given below is based on the characterization of compactness in terms of convergent subnets (Theorem 5.21) and is due to Paul Chernoff [**Che92**]. Similarly to certain standard results in functional analysis that also depend on the axiom of choice (e.g. the Hahn-Banach theorem), it uses the axiom in a somewhat indirect way, namely via *Zorn's lemma*, which is known to be equivalent to the axiom of choice. I do not want to go far enough into abstract set theory here to explain why it is equivalent: the proof is elementary but somewhat tedious, and you can find it explained e.g. in [**Jän05**] or [**Kel75**]. I would recommend reading through that proof exactly once in your life. For our purposes, we will just take the following statement of Zorn's lemma as a black box.

Lemma 6.4 (Zorn's lemma). *Suppose $(\mathcal{P},\prec)$ is a nonempty partially ordered set in which every totally ordered subset $\mathcal{A}\subset\mathcal{P}$ has an upper bound, i.e. for every subset in which all pairs $x,y\in\mathcal{A}$ satisfy $x\prec y$ or $y\prec x$, there exists an element $p\in\mathcal{P}$ such that $p\succ a$ for all $a\in\mathcal{A}$. Then every totally ordered subset $\mathcal{A}\subset\mathcal{P}$ also has an upper bound $p\in\mathcal{P}$ that is a maximal element, i.e. such that no $q\in\mathcal{P}$ with $q\neq p$ satisfies $q\succ p$.* ☐

Proof of Theorem 6.1, general case. We shall continue to denote elements of $\prod_{\alpha\in I}X_\alpha$ by functions $f: I\to\bigcup_{\alpha\in I}X_\alpha$ satisfying $f(\alpha)\in X_\alpha$ for each $\alpha\in I$. Assuming all the X_α are compact, it suffices by Theorem 5.21 to prove that every net $\{f_\beta\}_{\beta\in K}$ in $\prod_{\alpha\in I}X_\alpha$ has a cluster point. The idea of Chernoff's proof is as follows: we introduce below the notion of a "partial" cluster point, which may be a function defined only on a subset of I. We will show that the set of all partial cluster points has a partial order for which Zorn's lemma applies and delivers a maximal element. The last step is to show that a maximal element in the set of partial cluster points must in fact be a cluster point of $\{f_\beta\}_{\beta\in K}$.

To define partial cluster points, notice that for any subset $J\subset I$, restricting any function $f\in\prod_{\alpha\in I}X_\alpha$ to the smaller domain J defines an element $f|_J\in\prod_{\alpha\in J}X_\alpha$. We will refer to a pair (J,g) as a **partial cluster point** of the net $\{f_\beta\}_{\beta\in K}$ if J is a subset of I and $g\in\prod_{\alpha\in J}X_\alpha$ is a cluster point of the net $\{f_\beta|_J\}_{\beta\in K}$ in $\prod_{\alpha\in J}X_\alpha$ obtained by restricting the functions $f_\beta: I\to\bigcup_{\alpha\in I}X_\alpha$ to $J\subset I$. Let $\mathcal{P}$ denote the set of all partial cluster points of $\{f_\beta\}_{\beta\in K}$. It is easy to see that $\mathcal{P}$ is nonempty: indeed, for each individual $\alpha\in I$, the compactness of X_α implies that the net $\{f_\beta(\alpha)\}_{\beta\in K}$ in X_α has a cluster point $x_\alpha\in X_\alpha$, hence $(\{\alpha\},x_\alpha)\in\mathcal{P}$.

There is also an obvious partial order on $\mathcal{P}$: we shall write $(J,g)\leqslant(J',g')$ whenever $J\subset J'$ and $g=g'|_J$. In order to satisfy the main hypothesis of Zorn's lemma, we claim that every totally ordered subset $\mathcal{A}\subset\mathcal{P}$ has an upper bound. Being totally ordered means that for any two elements of $\mathcal{A}$, one is obtained from the other by restricting the function to a subset. We can therefore define a set $J_\infty\subset I$ with a function $g_\infty\in\prod_{\alpha\in J_\infty}X_\alpha$ by

$$J_\infty=\bigcup_{\{J\ |\ (J,g)\in\mathcal{A}\}}J,$$

with $g_\infty(\alpha)$ defined as $g(\alpha)$ for any $(J, g) \in \mathcal{A}$ such that $\alpha \in J$. The total ordering condition guarantees that (J_∞, g_∞) is independent of choices, but it is not immediately clear whether it is an element of $\mathcal{P}$, i.e. whether g_∞ is a cluster point of $\{f_\beta|_{J_\infty}\}_{\beta \in K}$. To see this, suppose $\mathcal{U} \subset \prod_{\alpha \in J_\infty} X_\alpha$ is a neighborhood of g_∞, and recall that by the definition of the product topology, this means

$$g_\infty \in \prod_{\alpha \in J_\infty} \mathcal{U}_\alpha \subset \mathcal{U}$$

for some collection of open sets $\mathcal{U}_\alpha \subset X_\alpha$ such that $\mathcal{U}_\alpha = X_\alpha$ for all α outside some finite subset $J_0 \subset J_\infty$. Since J_0 is finite, and $\mathcal{A}$ is totally ordered, there exists some $(J, g) \in \mathcal{A}$ such that $J_0 \subset J$. Then the fact that (J, g) is a partial cluster point means that for every $\beta_0 \in K$, there exists a $\beta > \beta_0$ for which

$$f_\beta|_J \in \prod_{\alpha \in J} \mathcal{U}_\alpha.$$

It follows that $f_\beta|_{J_\infty} \in \prod_{\alpha \in J_\infty} \mathcal{U}_\alpha$ as well, hence (J_∞, g_∞) is indeed a partial cluster point.

We can now apply Zorn's lemma and conclude that $\mathcal{P}$ has a maximal element $(J_M, g_M) \in \mathcal{P}$. We claim $J_M = I$, which means g_M is a cluster point of the original net $\{f_\beta\}_{\beta \in K}$ in $\prod_{\alpha \in I} X_\alpha$. Note that since $g_M \in \prod_{\alpha \in J_M} X_\alpha$ is a cluster point of $\{f_\beta|_{J_M}\}_{\beta \in K}$, Lemma 5.20 provides a subnet $\{f_{\phi(\gamma)}\}_{\gamma \in L}$ of $\{f_\beta\}_{\beta \in K}$ in $\prod_{\alpha \in I} X_\alpha$ whose restriction to J_M converges to g_M. But if $J_M \neq I$, then choosing an element $\alpha_0 \in I \backslash J_M$, we can exploit the fact that X_{α_0} is compact and use the same trick as in the proof of Tychonoff for finite products to find a further subnet that also converges at α_0 to some element $x_0 \in X_{\alpha_0}$. We have therefore found a subnet of $\{f_\beta\}_{\beta \in K}$ whose restriction to $J_M \cup \{\alpha_0\}$ converges to the function $g'_M \in \prod_{\alpha \in J_M \cup \{\alpha_0\}} X_\alpha$ defined by $g'_M|_{J_M} = g_M$ and $g'_M(\alpha_0) = x_0$. This means $(J_M \cup \{\alpha_0\}, g'_M) \in \mathcal{P}$ and $(J_M \cup \{\alpha_0\}, g'_M) > (J_M, g_M)$, which is a contradiction since (J_M, g_M) is maximal. □

EXERCISE 6.5. Consider the space $[0,1]^{\mathbb{R}}$ of all functions $f : \mathbb{R} \to [0,1]$, with the topology of pointwise convergence. Tychonoff's theorem implies that $[0,1]^{\mathbb{R}}$ is compact, but one can show that it is not first countable, so it need not be sequentially compact.

(a) For $x \in \mathbb{R}$ and $n \in \mathbb{N}$, let $x_{(n)} \in \{0, \dots, 9\}$ denote the nth digit to the right of the decimal point in the decimal expansion of x. Now define a sequence $f_n \in [0,1]^{\mathbb{R}}$ by setting $f_n(x) = \frac{x_{(n)}}{10}$. Show that for any subsequence f_{k_n} of f_n, there exists $x \in \mathbb{R}$ such that $f_{k_n}(x)$ does not converge, hence f_n has no pointwise convergent subsequence.
Food for thought: Could you do this if you also had to assume that x is rational? Presumably not, because $[0,1]^{\mathbb{Q}}$ is a product of countably many second countable spaces, and we've proved that such products are second countable (unlike $[0,1]^{\mathbb{R}}$). This implies that since $[0,1]^{\mathbb{Q}}$ is compact, it must also be sequentially compact.

(b) The compactness of $[0,1]^{\mathbb{R}}$ does imply that every sequence has a convergent *subnet*, or equivalently, a cluster point. Use this to deduce that for any given sequence $f_n \in [0,1]^{\mathbb{R}}$, there exists a function $f \in [0,1]^{\mathbb{R}}$ such that for every finite subset $X \subset \mathbb{R}$, some subsequence of f_n converges to f at all points in X.
Achtung: Pay careful attention to the order of quantifiers here. We're claiming that the element f exists independently of the finite set $X \subset \mathbb{R}$ on which we want some subsequence to converge to f. (If you could let f depend on the choice of subset X, this would be easy—but that is not allowed.) On the other hand, the actual choice of subsequence is allowed to depend on the subset X.

Challenge: Find a direct proof of the statement in part (b), without passing through Tychonoff's theorem. I do not know of any way to do this that isn't approximately as difficult as actually proving Tychonoff's theorem and dependent on the axiom of choice.

So much for Tychonoff's theorem. In truth, aside from the easy case of finite products, the general version of this theorem will probably not be mentioned again in this course. You may hear of it again if you take functional analysis since it lies in the background of the Banach-Alaoglu theorem on compactness in the *weak**-topology, and I will have occasion to mention it in *Topologie II* next semester in the context of the Eilenberg-Steenrod axioms for Čech homology. But right now we need to discuss a few more mundane things.

Topic 2: Separation axioms. Recall from Proposition 5.11 that closed subsets of compact spaces are always compact. Your intuition probably tells you that all compact sets are closed, but this in general is false. Here is a counterexample.

Example 6.6. Recall from Example 2.2 the so-called "line with two zeroes". We defined it as a quotient $X := (\mathbb{R} \times \{0,1\})/\sim$ by the equivalence relation such that $(x,0) \sim (x,1)$ for all $x \neq 0$, with a topology defined via the pseudometric $d([(x,i)],[(y,j)]) = |x-y|$, i.e. the open balls $B_r(x) := \{y \in X \mid d(y,x) < r\}$ for $x \in X$ and $r > 0$ form a base of the topology. Each $x \in \mathbb{R}\backslash\{0\}$ corresponds to a unique point $[(x,0)] = [(x,1)] \in X$, but for $x = 0$ there are two distinct points, which we shall abbreviate by

$$0_0 := [(0,0)] \in X \quad \text{and} \quad 0_1 := [(0,1)] \in X.$$

As we saw in Exercise 2.3, the one-point subset $\{0_1\} \subset X$ is not closed, but it certainly is compact since finite subsets are always compact (see Example 5.5). The failure of $\{0_1\}$ to be closed results from the fact that since $d(0_0, 0_1) = 0$, every neighborhood of 0_0 also contains 0_1, implying that $X\backslash\{0_1\}$ cannot be open.

The example of the line with two zeroes is pathological in various ways, e.g. it has the property that every sequence convergent to 0_1 also converges to the distinct point 0_0. We would now like to formulate some precise conditions to exclude such behavior. The most important of these will be the *Hausdorff* axiom, but there is a whole gradation of stronger or weaker variations on the same theme, known collectively as the **separation axioms** (*Trennungsaxiome*). Intuitively, they measure the degree to which topological notions such as convergence of sequences and continuity of maps can recognize the difference between two disjoint points or subsets.

Definition 6.7. A space X is said to satisfy axiom T_0 if for every pair of distinct points in X, there exists an open subset of X that contains one of these points but not the other.

Since almost all spaces we want to consider will satisfy the T_0 axiom, we should point out some examples of spaces that do not. One obvious example is any space of more than one element with the trivial topology: if the only open subset other than $\varnothing$ is X, then you clearly cannot find an open set that contains x and not $y \neq x$ or vice versa. A slightly more interesting example is the line with two zeroes as in Example 6.6 above, with the pseudometric topology: it fails to be a T_0 space because every open set that contains 0_0 or 0_1 must contain both of them.

Definition 6.8. A space X is said to satisfy **axiom** T_1 if for every pair of distinct points $x, y \in X$, there exist neighborhoods $\mathcal{U}_x \subset X$ of x and $\mathcal{U}_y \subset X$ of y such that $x \notin \mathcal{U}_y$ and $y \notin \mathcal{U}_x$.

Obviously every T_1 space is also T_0. The following alternative characterization of the T_1 axiom is immediate from the definitions:

Proposition 6.9. *A space X satisfies axiom T_1 if and only if for every point $x \in X$, the subset $\{x\} \subset X$ is closed.* □

Definition 6.10. A space X is said to satisfy **axiom** T_2 (the **Hausdorff** axiom) if for every pair of distinct points $x, y \in X$, there exist neighborhoods $\mathcal{U}_x \subset X$ of x and $\mathcal{U}_y \subset X$ of y such that $\mathcal{U}_x \cap \mathcal{U}_y = \varnothing$.

Every Hausdorff space is clearly also T_1 and T_0. Here is an easy criterion with which to recognize a non-Hausdorff space:

Exercise 6.11. Show that if X is Hausdorff, then for any sequence $x_n \in X$ satisfying $x_n \to x$ and $x_n \to y$, we have $x = y$.

Finding an example that is T_1 but not Hausdorff requires only a slight modification of our previous "line with two zeroes".

Example 6.12. Consider $X = (\mathbb{R} \times \{0,1\})/\sim$ again with $(x,0) \sim (x,1)$ for every $x \neq 0$, but instead of the pseudometric topology as in Example 6.6, assign it the quotient topology, meaning $\mathcal{U} \subset X$ is open if and only if its preimage under the projection map $\pi : \mathbb{R} \times \{0,1\} \to X : (x,i) \mapsto [(x,i)]$ is open. Recall that the quotient topology is the strongest topology for which π is a continuous map, and in this case, it turns out to be slightly stronger than the pseudometric topology. For example, the open set

$$\mathcal{V} := ((-1,1) \times \{0\}) \cup ((-1,0) \times \{1\}) \cup ((0,1) \times \{1\}) \subset \mathbb{R} \times \{0,1\}$$

is $\pi^{-1}(\mathcal{U})$ for $\mathcal{U} := \pi(\mathcal{V}) \subset X$, thus $\mathcal{U}$ is open in the quotient topology. But $\mathcal{U}$ contains 0_0 and not 0_1, so it is not an open set in the pseudometric topology. The existence of this set implies that X with the quotient topology satisfies T_0. By exchanging the roles of 0 and 1, one can similarly construct an open neighborhood of 0_1 that does not contain 0_0, so the space also satisfies T_1. But it does not satisfy T_2: even in the quotient topology, every neighborhood of 0_0 has nonempty intersection with every neighborhood of 0_1.

Exercise 6.11 has a converse of sorts, which I will state here only for first countable spaces. The countability axiom can be removed at the cost of talking about nets instead of sequences; I will leave the details of this as an exercise for the reader.

Proposition 6.13. *A first countable space X is Hausdorff if and only if the limit of every convergent sequence in X is unique.*

Proof. In light of Exercise 6.11, we just need to show that if X is a first countable space that is not Hausdorff, we can find a sequence $x_n \in X$ that converges to two distinct points $x, y \in X$. Since X is not Hausdorff, we can pick two distinct points x and y such that every neighborhood of x intersects every neighborhood of y. Fix countable neighborhood bases $X \supset \mathcal{U}_1 \supset \mathcal{U}_2 \supset \ldots \ni x$ and $X \supset \mathcal{V}_1 \supset \mathcal{V}_2 \ldots \ni y$. Then by assumption, for each $n \in \mathbb{N}$ there exists a point $x_n \in \mathcal{U}_n \cap \mathcal{V}_n$. It is now straightforward to verify that $x_n \to x$ and $x_n \to y$. □

The Hausdorff axiom can still be strengthened a bit by talking about neighborhoods of closed sets rather than points. This can be useful, for instance, when considering the quotient space X/A defined by collapsing some closed subset $A \subset X$ to a point; cf. Exercise 6.20 below.

Definition 6.14. A space X is called **regular** (*regulär*) if for every point $x \in X$ and every closed subset $A \subset X$ not containing x, there exist neighborhoods $\mathcal{U}_x \subset X$ of x and $\mathcal{U}_A \subset X$ of A such that $\mathcal{U}_x \cap \mathcal{U}_A = \emptyset$. We say X satisfies **axiom** T_3 if it is regular and also satisfies T_1.

Definition 6.15. A space X is called **normal** if for every pair of disjoint closed subsets $A, B \subset X$, there exist neighborhoods $\mathcal{U}_A \subset X$ of A and $\mathcal{U}_B \subset X$ of B such that $\mathcal{U}_A \cap \mathcal{U}_B = \emptyset$. We say X satisfies **axiom** T_4 if it is normal and also satisfies T_1.

Remark 6.16. The point of including T_1 in the definitions of T_3 and T_4 is that it makes each one-point subset $\{x\} \subset X$ closed, thus producing obvious implications

$$T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1 \Rightarrow T_0. \tag{6.1}$$

Without assuming T_1, it is possible for spaces to be regular or normal without being Hausdorff, though we will not consider any examples of this. In fact, almost all spaces we actually want to think about in this course will be Hausdorff, and most will also be normal, thus satisfying all of these axioms.

REMARK 6.17. Some of the above definitions, especially for axioms T_3 and T_4, can be found in a few not-quite-equivalent variations in various sources in the literature. One common variation is to interchange the meanings of "regular" with "T_3" and "normal" with "T_4", which destroys the first two implications in (6.1). These discrepancies are matters of convention which are to some extent arbitrary: you are free to choose your favorite convention, but must then be careful about stating your definitions precisely and remaining consistent.

We can now give a better answer to the question of when a compact set must also be closed.

THEOREM 6.18. *If X is Hausdorff, then every compact subset of X is closed.*

PROOF. Given a compact set $K \subset X$, we need to show that $X \backslash K$ is open, or equivalently, that every $x \in X \backslash K$ is contained in an open set disjoint from K. By assumption X is Hausdorff, so for each $y \in K$, we can find open neighborhoods $\mathcal{U}_y \subset X$ of x and $\mathcal{V}_y \subset X$ of y such that $\mathcal{U}_y \cap \mathcal{V}_y = \emptyset$. Then the sets $\{\mathcal{V}_y\}_{y \in K}$ form an open cover of K, and since the latter is compact by assumption, we obtain a finite subset $y_1, \ldots, y_N \in K$ such that

$$K \subset \mathcal{V}_{y_1} \cup \ldots \cup \mathcal{V}_{y_N}.$$

The set $\mathcal{U} := \mathcal{U}_{y_1} \cap \ldots \cap \mathcal{U}_{y_N}$ is then an open neighborhood of x and is disjoint from $\mathcal{V}_{y_1} \cup \ldots \cup \mathcal{V}_{y_N}$, implying in particular that it is disjoint from K. □

EXERCISE 6.19. Prove:

(a) A finite topological space satisfies the axiom T_1 if and only if it carries the discrete topology.

(b) X is a T_2 space (i.e. Hausdorff) if and only if the *diagonal* $\Delta := \{(x, x) \in X \times X\}$ is a closed subset of $X \times X$.

(c) Every compact Hausdorff space is regular, i.e. compact + $T_2 \Rightarrow T_3$.
Hint: The argument needed for this was already used in the proof of Theorem 6.18.

(d) Every metrizable space satisfies the axiom T_4 (in particular it is *normal*).
Hint: Given disjoint closed sets $A, A' \subset X$, each $x \in A$ admits a radius $\epsilon_x > 0$ such that the ball $B_{\epsilon_x}(x)$ is disjoint from A', and similarly for points in A' (why?). The unions of all these balls won't quite produce the disjoint neighborhoods you want, but try cutting their radii in half.

EXERCISE 6.20. Suppose X is a Hausdorff space and $\sim$ is an equivalence relation on X. Let $X/\!\sim$ denote the quotient space equipped with the quotient topology and denote by $\pi : X \to X/\!\sim$ the canonical projection. Given a subset $A \subset X$, we will sometimes also use the notation X/A explained in Exercise 5.16.

(a) A map $s : X/\!\sim \to X$ is called a **section** of π if $\pi \circ s$ is the identity map on $X/\!\sim$. Show that if a continuous section exists, then $X/\!\sim$ is Hausdorff.

(b) Show that if X is also regular and $A \subset X$ is a closed subset, then X/A is Hausdorff.

(c) Consider $X = \mathbb{R}$ with the non-closed subset $A = (0, 1]$. Which of the separation axioms $T_0, \ldots, T_4$ does X/A satisfy?

Just for fun: think about some other examples of Hausdorff spaces X with non-Hausdorff quotients $X/\!\sim$. What stops you from constructing continuous sections $X/\!\sim \to X$?

Remark 6.21. In earlier decades, it was common to define compactness slightly differently: what many papers and textbooks from the first half of the 20th centuary call a "compact space" is what we would call a "compact Hausdorff space". You should be aware of this discrepancy if you consult the older literature.

7. Connectedness and local compactness

We would like to formalize the idea that in some spaces, you can find a continuous path connecting any point to any other point, and in other spaces you cannot.

Definition 7.1. A space X is called **path-connected** (*wegzusammenhängend*) if for every pair of points $x, y \in X$, there exists a continuous map $\gamma : [0,1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

A subset of X is similarly called path-connected if it is a path-connected space in the subspace topology, which is equivalent to saying that any two points in the subset can be connected by a continuous path in that subset. We will refer to any maximal path-connected subset of a space X as a **path-component** (*Wegzusammenhangskomponente*) of X.

Exercise 7.2. Show that any two path-components of a space X must be either identical or disjoint, i.e. the path-components partition X into disjoint subsets. One can also express this by saying that there is a well-defined equivalence relation $\sim$ on X such that $x \sim y$ if and only if x and y belong to the same path-component. (Why is that an equivalence relation?)

The notion of path-connectedness is framed in terms of maps into X, but there is also a "dual" perspective based on functions defined on X. To motivate this, notice that if $f : X \to \{0,1\}$ is any continuous function and $x, y \in X$ belong to the same path-component, then continuity demands $f(x) = f(y)$. (We will formalize this observation in the proof of Theorem 7.13 below.)

Definition 7.3. A space X is **connected** (*zusammenhängend*) if every continuous map $X \to \{0,1\}$ is constant.

In many textbooks one finds a cosmetically different definition of connectedness in terms of subsets that are both open and closed, but the two definitions are equivalent due to the following result.

Proposition 7.4. *A space X is connected if and only if $\varnothing$ and X are the only subsets of X that are both open and closed.*

Proof. We prove first that the condition in this statement implies connectedness. The key observation is that the sets $\{0\}$ and $\{1\}$ in $\{0,1\}$ are each both open and closed, so if $f : X \to \{0,1\}$ is continuous, the same must hold for both $f^{-1}(0)$ and $f^{-1}(1)$ in X. Then one of these is the empty set and the other is X, so f is constant.

Conversely, suppose X contains a nonempty subset $X_0 \subset X$ that is both open and closed but $X_0 \neq X$. Then $X_1 := X \backslash X_0$ is also a nonempty open and closed subset, implying that X is the union of two disjoint open subsets X_0 and X_1. We can now define a nonconstant continuous function $f : X \to \{0,1\}$ by $f|_{X_0} = 0$ and $f|_{X_1} = 1$. Checking that it is continuous is easy since $\{0,1\}$ only contains four open sets: the main point is that $f^{-1}(0) = X_0$ and $f^{-1}(1) = X_1$ are both open. □

Remark 7.5. The important fact about $\{0,1\}$ used in the above proof was that it is a space of more than one element with the discrete topology: officially $\{0,1\}$ carries the subspace topology as a subset of $\mathbb{R}$, but this happens to match the discrete topology since 0 and 1 are each centers of open balls in $\mathbb{R}$ that do not touch any other points of $\{0,1\}$. If we preferred, we could have replaced Definition 7.3 with the condition that every continuous map $f : X \to Y$ to any space Y with the discrete topology is constant.

We can of course also talk about **connected subsets** $A \subset X$, meaning subsets that become connected spaces with the subspace topology. Spaces or subsets that are not connected are sometimes called **disconnected**. By analogy with path-components, any maximal connected subset of X will be called a **connected component** (*Zusammenhangskomponente*) of X.

Proposition 7.6. *Any two connected components $A, B \subset X$ are either identical or disjoint.*

Proof. If A and B are both maximal connected subsets of X and $A \cap B \neq \varnothing$, then we claim that $A \cup B$ is also connected. Indeed, any continuous function $f : A \cup B \to \{0, 1\}$ must restrict to constant functions on both A and B, so if $y \in A \cap B$, then $f(x) = f(y)$ for every $x \in A \cup B$, implying that every continous function $A \cup B \to \{0, 1\}$ is constant. Now if A and B are not identical, then the set $A \cup B$ is strictly larger than either A or B, giving a contradiction to the maximality assumption. ☐

Example 7.7. For any collection $\{X_\alpha\}_{\alpha \in I}$ of connected spaces, the disjoint union $X := \coprod_{\alpha \in I} X_\alpha$ has the individual spaces $X_\alpha \subset X$ for $\alpha \in I$ as its connected components. Indeed, endowing X with the disjoint union topology makes each of the subsets $X_\alpha \subset X$ open, and since $X \backslash X_\alpha = \bigcup_{\beta \neq \alpha} X_\beta$ is then also open, it follows that X_α is also closed. Any strictly larger set $A \subset X$ with $X_\alpha \subset A$ could not then be connected, as it would contain X_α as a nonempty proper open and closed subset; this makes X_α a *maximal* connected subset of X.

Exercise 7.8. Show that if the spaces X_α in Example 7.7 are also path-connected, then they also form the path-components of the disjoint union $X = \coprod_{\alpha \in I} X_\alpha$.

For an arbitrary space X, let us choose an index set I with which to label each connected component of X, so the connected components from a collection of spaces $\{X_\alpha\}_{\alpha \in I}$, each of which is a subset $X_\alpha \subset X$ endowed with the subspace topology. Proposition 7.6 shows that $X_\alpha \cap X_\beta = \varnothing$ whenever $\alpha \neq \beta$, and obviously $\bigcup_{\alpha \in I} X_\alpha = X$, so as sets, there is a canonical bijective correspondence between X and the disjoint union $\coprod_{\alpha \in I} X_\alpha$. It is natural to wonder: is this correspondence a homeomorphism? It is easy to see that it is continuous in at least one direction: the individual subsets $X_\alpha \subset X$ come with inclusion maps $i_\alpha : X_\alpha \hookrightarrow X$, and endowing X_α with the subspace topology makes i_α continuous. The canonical bijection from $\coprod_{\alpha \in I} X_\alpha$ to X can then be written as

$$\coprod_{\alpha \in I} i_\alpha : \coprod_{\alpha \in I} X_\alpha \to X, \tag{7.1}$$

meaning it is the unique map whose restriction to each of the subsets $X_\alpha \subset \coprod_{\beta \in I} X_\beta$ is precisely i_α. The definition of the disjoint union topology makes this map automatically continuous. The following example shows however that, in general, its inverse need not be continuous.

Example 7.9. The set $\mathbb{Q}$ of rational numbers is a perfectly nice algebraic object, but when endowed with the subspace topology as a subset of $\mathbb{R}$, it becomes a very badly behaved topological space. We claim that if $A \subset \mathbb{Q}$ is any subset with more than one element, then A is disconnected. Indeed, given $x, y \in A$ with $x < y$, we can find an irrational number $r \in \mathbb{R} \backslash \mathbb{Q}$ with $x < r < y$, and the sets $A_- := A \cap (-\infty, r)$ and $A_+ := A \cap (r, \infty)$ are then nonempty open subsets of A which are complements of each other, hence both are open and closed. This proves that the connected components of $\mathbb{Q}$ are simply the one-point subspaces $\{x\} \subset \mathbb{Q}$ for all $x \in \mathbb{Q}$, so the map (7.1) in this case takes the form

$$\coprod_{x \in \mathbb{Q}} \{x\} \to \mathbb{Q}.$$

The domain and target of this map are the same set, and the map itself is the identity, but the two sets are endowed with very different topologies: in particular, the domain carries the discrete topology, while $\mathbb{Q}$ on the right hand side carries the subspace topology that it inherits from the standard topology of $\mathbb{R}$. The identity map is thus continuous—indeed, every map defined on a space with the discrete topology is continuous—but it is not a homeomorphism, because the discrete topology contains many open sets that are not open in the standard topology of $\mathbb{Q}$.

Example 7.9 shows that while every space X has a natural bijective correspondence with the disjont union $\coprod_{\alpha \in I} X_\alpha$ of its connected components, the natural topology on $\coprod_{\alpha \in I} X_\alpha$ may in general be different from the original topology of X. We've seen for instance that each individual X_α is automatically both an open and closed subset of $\coprod_{\beta \in I} X_\beta$, thus there is no hope of (7.1) being a homeomorphism unless X_α is also an open and closed subset of X. The example of $\mathbb{Q}$ shows that the latter is not always true: the 1-point connected components $\{x\} \subset \mathbb{Q}$ are closed subsets, but they are not open. The fact that they are closed turns out to be a completely general phenomenon:

Proposition 7.10. *Every connected component $A \subset X$ of a space X is a closed subset.*

Proof. Assume $A \subset X$ is a maximal connected subset. Recall from Definition 3.1 that the closure $\bar{A} \subset X$ of A is the set of all points $x \in X$ for which every neighborhood of x intersects A. If we equip $\bar{A}$ with the subspace topology and view it as a topological space in itself, with $A \subset \bar{A}$ as a subset, then the closure of A in $\bar{A}$ is still $\bar{A}$: indeed, every neighborhood in $\bar{A}$ of a point $x \in \bar{A}$ takes the form $\mathcal{U} \cap \bar{A}$ for some neighborhood $\mathcal{U}$ of x in X, implying that $\mathcal{U}$ intersects A, and therefore so does $\mathcal{U} \cap \bar{A}$.

Now suppose $f : \bar{A} \to \{0, 1\}$ is a continuous function. Its restriction to A is then also continuous, and therefore constant, since A is connected; let us write $f(A) = \{i\} \subset \{0, 1\}$. Then since $\{i\}$ is a closed subset of $\{0, 1\}$ and f is continuous, $f^{-1}(i)$ is a closed subset of $\bar{A}$ that contains A, and it therefore also contains the closure $\bar{A}$. This implies that f is in fact constant on $\bar{A}$, and thus proves that $\bar{A}$ is connected. Since A is a maximal connected subset, we conclude $A = \bar{A}$, meaning A is closed. □

We note one obvious case in which connected components will necessarily be both closed and open: here openness follows from the fact that the complement of a connected component is a union of disjoint connected components, and finite unions of closed sets are closed.

Corollary 7.11. *If X is a space with only finitely many connected components, then each of them is both closed and open.* □

Exercise 7.12. If $\{X_\alpha \subset X\}_{\alpha \in I}$ are the connected components of a space X, show that the canonical continuous bijection (7.1) from $\coprod_{\alpha \in I} X_\alpha$ to X is a homeomorphism if and only if every X_α is an open subset of X. (In particular, Corollary 7.11 implies that this is always true if I is finite, and we will see in Prop. 7.18 below that it is also true if X is locally connected.)

It is time to clarify the relationship between connectedness and path-connectedness.

Theorem 7.13. *Every path-connected space X is connected.*

Proof. If X is not connected, then there exist points $x, y \in X$ and a continuous function $f : X \to \{0, 1\}$ such that $f(x) = 0$ and $f(y) = 1$. But if X is path-connected, then there also exists a continuous map $\gamma : [0, 1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$. The composition $g := f \circ \gamma$ is then a continuous function $g : [0, 1] \to \{0, 1\}$ satisfying $g(0) = 0$ and $g(1) = 1$, and this violates the intermediate value theorem. □

Surprisingly, the converse of this theorem is false.

Example 7.14. Define $X \subset \mathbb{R}^2$ to be the subset of $\mathbb{R}^2$ consisting of the vertical line $\{x = 0\}$ and the graph of the equation $\{y = \sin(1/x)\}$ for $x \neq 0$. The latter is a sine curve that oscillates more and more rapidly as $x \to 0$. We claim that

$$X_0 := \{x = 0\}$$

is a path-component of X. It clearly is path-connected, so we need to show that there does not exist any continuous path $\gamma : [0, 1] \to X$ that begins on the sine curve $\{y = \sin(1/x)\}$ and ends on the line $\{x = 0\}$. Since $\{x = 0\}$ is a closed subset, the preimage of this set under γ is closed (and therefore compact) in $[0, 1]$, implying that it has a minimum $\tau \in (0, 1]$. We can therefore restrict our path to $\gamma : [0, \tau] \to X$ and assume that it lies on the sine curve for all $0 \leqslant t < \tau$ but ends on the vertical line at $t = \tau$. Now observe that due to the rapid oscillation as $x \to 0$, we can find for any $y \in [-1, 1]$ a sequence $t_n \in [0, \tau)$ with $t_n \to \tau$ such that $\gamma(t_n) \to (0, y)$. The point y here is arbitrary, yet continuity of γ requires $\gamma(t_n) \to \gamma(\tau)$, so this is a contradiction and proves the claim. In particular, this proves that X is not path-connected. The other path-components of X are now easy to identify: they are

$$X_- := X \cap \{x < 0\} \quad \text{and} \quad X_+ := X \cap \{x > 0\},$$

the portions of the sine curve lying to the left and right of X_0, so there are three path-components in total. The path-components are path-connected and therefore (by Theorem 7.13) also connected. But neither X_- nor X_+ is closed, so by Prop. 7.10, neither of these can be a connected component. The maximal connected subset containing X_-, for instance, must be a closed set containing X_- and therefore contains the closure $\overline{X_-}$, which includes points in X_0. Since X_0 is path-connected, it follows that the connected component containing X_- also contains all of X_0. But the same argument applies equally well to X_+, and these two observations together imply that all three path-components are in the same connected component, i.e. X is connected.

The space in Example 7.14 is sometimes called the *topologist's sine curve*. There is a certain "local" character to the pathologies of this space, i.e. part of the reason for its bizarre properties is that one can zoom in on certain points in X arbitrarily far without making it look more reasonable—in particular this is true for the points in X_0 that are in the closure of X_- and X_+. One can use neighborhoods of points to formalize this notion of "zooming in" arbitrarily far.

Definition 7.15. A space X is **locally connected** (*lokal zusammenhängend*) if for all points $x \in X$, every neighborhood of x contains a connected neighborhood of x.

The version of this for path-connectedness is completely analogous.

Definition 7.16. A space X is **locally path-connected** (*lokal wegzusammenhängend*) if for all points $x \in X$, every neighborhood of x contains a path-connected neighborhood of x.

Local path-connectedness obviously implies local connectedness by Theorem 7.13. Since most spaces we can easily imagine will have both properties, it is important at this juncture to look at some examples that do not. The topologist's sine curve in Example 7.14 is one such space: it is not locally connected (even though it is connected), since sufficiently small neighborhoods of points $(0, y) \in X$ for $-1 < y < 1$ always have infinitely many pieces of the sine curve passing through and are thus disconnected. Here is an example that is path-connected, but not locally:

Example 7.17. Let $X \subset \mathbb{R}^2$ denote the compact set

$$X = \left(\bigcup_{n=1}^{\infty} L_n \right) \cup L_\infty,$$

where for each $n \in \mathbb{N}$, L_n denotes the straight line segment from $(0,1)$ to $(1/n, 0)$, and the case $n = \infty$ is included for the vertical segment from $(0,1)$ to $(0,0)$. Then sufficiently small neighborhoods of $(0,0)$ in this space are never connected, so X is not locally connected. Notice however that there are continuous paths along the line segments L_n from any point in X to $(0,1)$, so X is path-connected.

Proposition 7.18. *If X is locally connected, then its connected components are open subsets. Similarly, if X is locally path-connected, then its path-components are open subsets.*

Proof. If X is locally connected and $A \subset X$ is a maximal connected subset, then for each $x \in A$, fix a connected neighborhood $\mathcal{U}_x \subset X$ of x. Now for $\mathcal{U} := \bigcup_{x \in A} \mathcal{U}_x$, any continuous function $f : \mathcal{U} \to \{0,1\}$ must restrict to a constant on each $\mathcal{U}_x$ and also on A, implying that f is constant, hence $\mathcal{U}$ is connected. The maximality of A thus implies $A = \mathcal{U}$, but $\mathcal{U}$ is also a neighborhood of A and thus contains an open set containing A, therefore A is open.

A completely analogous argument works in the locally path-connected case, taking path-connected neighborhoods $\mathcal{U}_x$ and using the fact that their union must also be path-connected. ☐

A consequence of this result is that the phenomenon allowing certain spaces to be connected but not path-connected is essentially local:

Theorem 7.19. *Every space that is connected and locally path-connected is also path-connected.*

Proof. If X is locally path-connected, then by Prop. 7.18 its path-components are open. Then if $A \subset X$ is a path-component, $X \backslash A$ is a union of path-components and is therefore also open, implying that A is both open and closed. If X is connected, it follows that $A = X$, so X is a path-component. ☐

Exercise 7.20. In this exercise we show that products of (path-)connected spaces are also (path-)connected, so long as one uses the correct topology on the product.

(a) Prove that if X and Y are both connected, then so is $X \times Y$.
Hint: Start by showing that for any $x \in X$ and $y \in Y$, the subsets $\{x\} \times Y$ and $X \times \{y\}$ in $X \times Y$ are connected. Then think about continuous maps $X \times Y \to \{0,1\}$.

(b) Show that for any collection of path-connected spaces $\{X_\alpha\}_{\alpha \in I}$, the space $\prod_{\alpha \in I} X_\alpha$ is path-connected in the usual product topology.
Hint: You might find Exercise 4.5 helpful.

(c) Consider $\mathbb{R}^{\mathbb{N}}$ with the "box topology" which we discussed in Exercise 4.6. Show that the set of all elements $f \in \mathbb{R}^{\mathbb{N}}$ represented as functions $f : \mathbb{N} \to \mathbb{R}$ that satisfy $\lim_{n \to \infty} f(n) = 0$ is both open and closed, hence $\mathbb{R}^{\mathbb{N}}$ in the box topology is not connected (and therefore also not path-connected).

The rest of this exercise is aimed at generalizing part (a) to the statement that for an arbitrary collection $\{X_\alpha\}_{\alpha \in I}$ of connected (but not necessarily path-connected) spaces, $\prod_{\alpha \in I} X_\alpha$ with the product topology is also connected. Choose a point $\{c_\alpha\}_{\alpha \in I} \in \prod_{\alpha \in I} X_\alpha$ and, for each finite subset $J \subset I$ of the index set, consider the set

$$X_J := \left\{ \{x_\alpha\}_{\alpha \in I} \in \prod_{\alpha \in I} X_\alpha \;\middle|\; x_\beta = c_\beta \text{ for all } \beta \in I \backslash J \right\},$$

endowed with the subspace topology that it inherits from the product topology of $\prod_{\alpha \in I} X_\alpha$.

(d) Show that for every choice of finite subset $J \subset I$, X_J is connected.
Hint: This is not really that different from part (a).

(e) Deduce that the union $\bigcup_J X_J \subset \prod_{\alpha \in I} X_\alpha$ is also connected, where J ranges over the set of all finite subsets of I.

(f) Show that the closure of the subset $\bigcup_J X_J \subset \prod_{\alpha\in I} X_\alpha$ is $\prod_{\alpha\in I} X_\alpha$, and deduce that $\prod_{\alpha\in I} X_\alpha$ is also connected.

With the definition of local connectedness in mind, we now briefly revisit the subject of compactness.

Definition 7.21. A space X is **locally compact** (*lokal kompakt*) if every point $x \in X$ has a compact neighorhood.

Local compactness is one of the notions for which one can find multiple inequivalent definitions in the literature, but as we'll see in a moment, all the plausible definitions of this concept are equivalent if we only consider Hausdorff spaces. Let's first note a few examples.

Example 7.22. The Euclidean space $\mathbb{R}^n$ is locally compact, and more generally, so is any closed subset $X \subset \mathbb{R}^n$ endowed with the subspace topology. Indeed, since closed and bounded subsets of $\mathbb{R}^n$ are compact, every $x \in X \subset \mathbb{R}^n$ has a compact neighborhood of the form $\overline{B_r(x)} \cap X$ for any $r > 0$.

Example 7.23. This is a non-example: a Hilbert space is not locally compact if it is infinite dimensional. This is due to the fact that every neighborhood of a point x must contain some closed ball $\overline{B_r(x)}$, but the latter is not compact (cf. Remark 5.8).

Example 7.24. Since a space is a neighborhood of all of its points, every compact space is (trivially) locally compact.

The last example is the one that becomes slightly controversial if you look at alternative definitions of local compactness in the literature, and indeed, if we had phrased Definition 7.21 more analogously to the definition of local (path-)connectedness, it would be easy to imagine spaces that are compact without being locally compact. As it happens, this never happens for Hausdorff spaces, and since we will mainly be interested in Hausdorff spaces, we shall take the following result as an excuse to avoid worrying any further about discrepancies in definitions. It will also be a useful result in its own right.

Theorem 7.25. *If X is Hausdorff, then the following conditions are equivalent:*

(i) X is locally compact (in the sense of Definition 7.21);
(ii) For all $x \in X$, every neighborhood of x contains a compact neighborhood of x;
(iii) If $K \subset \mathcal{U} \subset X$ where K is compact and $\mathcal{U}$ is open, then $K \subset \mathcal{V} \subset \bar{\mathcal{V}} \subset \mathcal{U}$ for some open set $\mathcal{V}$ with compact closure $\bar{\mathcal{V}}$.

Proof. Since single point subsets $\{x\} \subset X$ are always compact, it is clear that (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The implication (ii) $\Rightarrow$ (iii) is a relatively straightforward exercise using the finite covering property for the compact set K. We will therefore focus on the implication (i) $\Rightarrow$ (ii).

Assume we are given a neighborhood $\mathcal{U} \subset X$ of x and would like to find a compact neighborhood inside $\mathcal{U}$. By assumption, x also has a compact neighborhood $K \subset X$. It will do no harm to replace $\mathcal{U}$ with a smaller neighorhood such as the interior of $\mathcal{U} \cap K$, so without loss of generality, let us assume $\mathcal{U}$ is open and contained in K, in which case (since X is Hausdorff and K is therefore closed) its closure $\bar{\mathcal{U}}$ is also contained in K and is thus compact. We define the *boundary* of $\bar{\mathcal{U}}$ by

$$\partial\bar{\mathcal{U}} = \bar{\mathcal{U}} \cap \overline{X\backslash\mathcal{U}}.$$

This is a closed subset of $\bar{\mathcal{U}}$ and is therefore also compact, and we observe that since x is contained in a neighborhood disjoint from $X\backslash\mathcal{U}$, x is not in the closure $\overline{X\backslash\mathcal{U}}$ and thus

$$x \notin \partial\bar{\mathcal{U}}.$$

Since X is Hausdorff, for every $y \in \partial\bar{\mathcal{U}}$ there exists a pair of open neighborhoods

$$x \in A_y \subset X, \quad y \in B_y \subset X \quad \text{such that} \quad A_y \cap B_y = \emptyset.$$

Then the sets B_y for $y \in \partial\bar{\mathcal{U}}$ form an open cover of the compact set $\partial\bar{\mathcal{U}}$, hence there exists a finite subset $\{y_1, \ldots, y_N\} \subset \partial\bar{\mathcal{U}}$ such that

$$\partial\bar{\mathcal{U}} \subset \bigcup_{i=1}^{N} B_{y_i}.$$

Now the set

$$\mathcal{V} := \mathcal{U} \cap \left(\bigcap_{i=1}^{N} A_{y_i} \right)$$

is an open neighborhood of x contained in $\mathcal{U}$ and disjoint from the neighborhood $\bigcup_{i=1}^{N} B_{y_i}$ of $\partial\bar{\mathcal{U}}$. The latter implies that for any $y \in \partial\bar{\mathcal{U}}$, y has a neighborhood disjoint from $\mathcal{V}$, hence $y \notin \bar{\mathcal{V}}$. Similarly, $\mathcal{V} \subset \mathcal{U}$ implies y cannot be in the closure of $\mathcal{V}$ if it is in the interior of $\overline{X \backslash \mathcal{U}}$, so we conclude $\bar{\mathcal{V}} \subset \mathcal{U}$. The compactness of $\bar{\mathcal{V}}$ follows because it is a closed subset of $\bar{\mathcal{U}}$ and the latter is compact. ☐

Exercise 7.26. Prove the implication that was skipped in the proof of Theorem 7.25 above, namely: if X is locally compact and Hausdorff, then for any nested pair of subsets $K \subset \mathcal{U} \subset X$ with K compact and $\mathcal{U}$ open, there exists an open set $\mathcal{V} \subset X$ with compact closure $\bar{\mathcal{V}}$ such that $K \subset \mathcal{V} \subset \bar{\mathcal{V}} \subset \mathcal{U}$.

Exercise 7.27. There is a cheap trick to view any topological space as a compact space with a single point removed. For a space X with topology $\mathcal{T}$, let $\{\infty\}$ denote a set consisting of one element that is not in X, and define the **one point compactification** of X as the set $X^* = X \cup \{\infty\}$ with topology $\mathcal{T}^*$ consisting of all subsets in $\mathcal{T}$ plus all subsets of the form $(X \backslash K) \cup \{\infty\} \subset X^*$ where $K \subset X$ is closed and compact.

(a) Verify that $\mathcal{T}^*$ is a topology and that X^* is always compact.
(b) Show that if X is first countable and Hausdorff, a sequence in $X \subset X^*$ converges to $\infty \in X^*$ if and only if it has no convergent subsequence with a limit in X. Conclude that if X is first countable and Hausdorff, X^* is sequentially compact.
(c) Show that for $X = \mathbb{R}$, X^* is homeomorphic to S^1. *(More generally, one can use stereographic projection to show that the one point compactification of $\mathbb{R}^n$ is homeomorphic to S^n.)*
(d) Show that if X is already compact, then X^* is homeomorphic to the disjoint union $X \amalg \{\infty\}$.
(e) Show that X^* is Hausdorff if and only if X is both Hausdorff and locally compact.

Notice that $\mathbb{Q}$ is not locally compact, since every neighborhood of a point $x \in \mathbb{Q}$ contains sequences without convergent subsequences, e.g. any sequence of rational numbers that converges to an irrational number sufficiently close to x. The one point compactification $\mathbb{Q}^*$ is a compact space, and by part (b) it is also sequentially compact, but those are practically the only nice things we can say about it.

(f) Show that for any $x \in \mathbb{Q}$, every neighborhood of x in $\mathbb{Q}^*$ intersects every neighborhood of ∞, so in particular, $\mathbb{Q}^*$ is not Hausdorff.
Advice: Do not try to argue in terms of sequences with non-unique limits (cf. part (g) below), and do not try to describe precisely what arbitrary compact subsets of $\mathbb{Q}$ can look like (the answer is not nice). One useful thing you can say about arbitrary compact subsets of $\mathbb{Q}$ is that they can never contain the intersection of $\mathbb{Q}$ with any open interval. (Why not?)

(g) Show that every convergent sequence in $\mathbb{Q}^*$ has a unique limit. *(Since $\mathbb{Q}^*$ is not Hausdorff, this implies via Proposition 6.13 that $\mathbb{Q}^*$ is not first countable—in particular, ∞ does not have a countable neighborhood base.)*

(h) Find a point in $\mathbb{Q}^*$ with a neighborhood that does not contain any compact neighborhood.

Exercise 7.28. Given spaces X and Y, let $C(X,Y)$ denote the set of all continuous maps from X to Y, and consider the natural *evaluation map*

$$\mathrm{ev} : C(X,Y) \times X \to Y : (f,x) \mapsto f(x).$$

It is easy to show that ev is a continuous map if we assign the discrete topology to $C(X,Y)$, but usually one can also find more interesting topologies on $C(X,Y)$ for which ev is continuous. The **compact-open topology** is defined via a subbase consisting of all subsets of the form

$$\mathcal{U}_{K,V} := \left\{ f \in C(X,Y) \;\middle|\; f(K) \subset V \right\},$$

where K ranges over all compact subsets of X, and V ranges over all open subsets of Y. Prove:

(a) If Y is a metric space, then convergence of a sequence $f_n \in C(X,Y)$ in the compact-open topology means that f_n converges uniformly on all compact subsets of X.

(b) If $C(X,Y)$ carries the topology of pointwise convergence (i.e. the subspace topology defined via the obvious inclusion $C(X,Y) \subset Y^X$), then ev is not sequentially continuous in general.

(c) If $C(X,Y)$ carries the compact-open topology, then ev is always sequentially continuous.

(d) If $C(X,Y)$ carries the compact-open topology and X is locally compact and Hausdorff, then ev is continuous.

(e) Every topology on $C(X,Y)$ for which ev is continuous contains the compact-open topology. (This proves that if X is locally compact and Hausdorff, the compact-open topology is the weakest topology for which the evaluation map is continuous.)
Hint: If $(f_0,x_0) \in \mathrm{ev}^{-1}(V)$ where $V \subset Y$ is open, then $(f_0,x_0) \in \mathcal{O} \times \mathcal{U} \subset \mathrm{ev}^{-1}(V)$ for some open $\mathcal{O} \subset C(X,Y)$ and $\mathcal{U} \subset X$. Is $\mathcal{U}_{K,V}$ a union of sets $\mathcal{O}$ that arise in this way?

(f) For the compact-open topology on $C(\mathbb{Q},\mathbb{R})$, $\mathrm{ev} : C(\mathbb{Q},\mathbb{R}) \times \mathbb{Q} \to \mathbb{R}$ is not continuous.

Exercise 7.29. One of the good reasons to use the notation X^Y for the set of all functions $f : Y \to X$ between two sets is that there is an obvious bijection

$$Z^{X\times Y} \to (Z^Y)^X$$

sending a function $F : X \times Y \to Z$ to the function $\Phi : X \to Z^Y$ defined by

$$\Phi(x)(y) = F(x,y). \tag{7.2}$$

The existence of this bijection is sometimes called the *exponential law* for sets. In this exercise we will explore to what extent the exponential law carries over to topological spaces and continuous maps. We will see that this is also related to the question of how to define a natural topology on the group of homeomorphisms of a space.

If X and Y are topological spaces, let us denote by $C(X,Y)$ the space of all continuous maps $X \to Y$, with the compact-open topology, which has a subbase consisting of all sets of the form

$$\mathcal{U}_{K,V} := \left\{ f \in C(X,Y) \;\middle|\; f(K) \subset V \right\}$$

for $K \subset X$ compact and $V \subset Y$ open (see Exercise 7.28 above). Assume Z is also a topological space.

(a) Prove that if $F : X \times Y \to Z$ is continuous, then the correspondence (7.2) defines a continuous map $\Phi : X \to C(Y,Z)$.

(b) Prove that if Y is locally compact and Hausdorff, then the converse also holds: any continuous map $\Phi : X \to C(Y,Z)$ defines a continuous map $F : X \times Y \to Z$ via (7.2).

Let's pause for a moment to observe what these two results imply for the case $X := I = [0, 1]$. First, here is a quick definition of a notion that will appear very often in the remainder of this course: given two continuous maps $f_0, f_1 : Y \to Z$, a continuous map

$$h : I \times Y \to Z \qquad \text{such that} \qquad h(0, \cdot) = f_0 \text{ and } h(1, \cdot) = f_1$$

is called a **homotopy** (*Homotopie*) between f_0 and f_1, and we call f_0 and f_1 **homotopic** (*homotop*) if a homotopy between them exists. According to part (a), a homotopy between two maps $Y \to Z$ can always be regarded as a continuous path in $C(Y, Z)$, and part (b) says that the converse is also true if Y is locally compact and Hausdorff, hence two maps $Y \to Z$ are homotopic if and only if they lie in the same path-component of $C(Y, Z)$.[5]

(c) Deduce from part (b) a new proof of the following result from Exercise 7.28(d): if X is locally compact and Hausdorff, then the *evaluation map* $\mathrm{ev} : C(X, Y) \times X \to Y : (f, x) \mapsto f(x)$ is continuous.
Hint: This is very easy if you look at it from the right perspective.
Remark: If you were curious to see a counterexample to part (b) in a case where Y is not locally compact, you could now extract one from Exercise 7.28(f).

(d) The following cannot be deduced directly from part (b), but it is a similar result and requires a similar proof: show that if Y is locally compact and Hausdorff, then

$$C(X, Y) \times C(Y, Z) \to C(X, Z) : (f, g) \mapsto g \circ f$$

is a continuous map.
Hint: Exercise 7.26 is useful here.

Now let's focus on maps from a space X to itself. A group G with a topology is called a **topological group** if the maps

$$G \times G \to G : (g, h) \mapsto gh \qquad \text{and} \qquad G \to G : g \mapsto g^{-1}$$

are both continuous. Common examples include the standard matrix groups $\mathrm{GL}(n, \mathbb{R})$, $\mathrm{GL}(n, \mathbb{C})$ and their subgroups, which have natural topologies as subsets of the vector space of (real or complex) n-by-n matrices. Another natural example to consider is the group

$$\mathrm{Homeo}(X) = \left\{ f \in C(X, X) \;\middle|\; f \text{ is bijective and } f^{-1} \in C(X, X) \right\}$$

for any topological space X, where the group operation is defined via composition of maps. We would like to know what topologies can be assigned to $C(X, X)$ so that $\mathrm{Homeo}(X) \subset C(X, X)$, with the subspace topology, becomes a topological group. Notice that the discrete topology clearly works; this is immediate because all maps between spaces with the discrete topology are automatically continuous, so there is nothing to check. But the discrete topology is not very interesting. Let $\mathcal{T}_H$ denote the topology on $C(X, X)$ with subbase consisting of all sets of the form $\mathcal{U}_{K,V}$ and $\mathcal{U}_{X \setminus V, X \setminus K}$, where again $K \subset X$ can be any compact subset and $V \subset X$ any open subset. Notice that if X is compact and Hausdorff, then for any V open and K compact, $X \setminus V$ is compact and $X \setminus K$ is open, thus $\mathcal{T}_H$ is again simply the compact-open topology. But if X is not compact or Hausdorff, $\mathcal{T}_H$ may be stronger than the compact-open topology.

---

[5] Since $C(X \times Y, Z)$ and $C(X, C(Y, Z))$ both have natural topologies in terms of the compact-open topology, you may be wondering whether the correspondence (7.2) defines a homeomorphism between them. The answer to this is more complicated than one would like, but Steenrod showed in a famous paper in 1967 [Ste67] that the answer is "yes" if one restricts attention to spaces that are *compactly generated*, a property that most respectable spaces have. The caveat is that $C(X, Y)$ in the compact-open topology will not always be compactly generated if X and Y are, so one must replace the compact-open topology by a slightly stronger one that is compactly generated but otherwise has the same properties for most practical purposes. If you want to know what "compactly generated" means and why it is a useful notion, see [Ste67]. These issues are somewhat important in homotopy theory at more advanced levels, though it is conventional to worry about them as little as possible.

(e) Show that if X is locally compact and Hausdorff, then $\mathrm{Homeo}(X)$ with the topology $\mathcal{T}_H$ is a topological group.
Hint: Notice that $f(K) \subset V$ *if and only if* $f^{-1}(X \backslash V) \subset X \backslash K$. *Use this to show directly that* $f \mapsto f^{-1}$ *is continuous, and reduce the rest to what was proved already in part (d).*

Conclusion: We've shown that if X *is compact and Hausdorff, then* $\mathrm{Homeo}(X)$ *with the compact-open topology is a topological group. This is actually true under somewhat weaker hypotheses, e.g. it suffices to know that* X *is Hausdorff, locally compact and locally connected. (If you're interested, a quite clever proof of this fact may be found in* [**Are46**].*)*

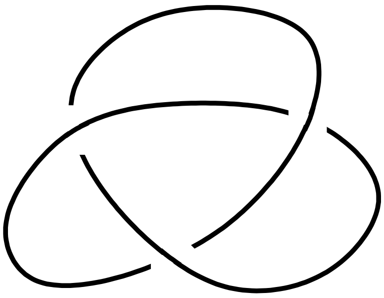
Just for fun, here's an example to show that just being locally compact and Hausdorff is not enough: let $X = \{0\} \cup \{e^n \mid n \in \mathbb{Z}\} \subset \mathbb{R}$ *with the subspace topology, and notice that* X *is neither compact (since it is unbounded) nor locally connected (since every neighborhood of* 0 *is disconnected). Consider the sequence* $f_k \in \mathrm{Homeo}(X)$ *defined for* $k \in \mathbb{N}$ *by* $f_k(0) = 0$, $f_k(e^n) = e^{n-1}$ *for* $n \leqslant -k$ *or* $n > k$, $f_k(e^n) = e^n$ *for* $-k < n < k$, *and* $f_k(e^k) = e^{-k}$. *It is not hard to show that in the compact-open topology on* $C(X, X)$, $f_k \to \mathrm{Id}$ *but* $f_k^{-1} \not\to \mathrm{Id}$ *as* $k \to \infty$, *hence the map* $\mathrm{Homeo}(X) \to \mathrm{Homeo}(X) : f \mapsto f^{-1}$ *is not continuous.*

8. Paths, homotopy and the fundamental group

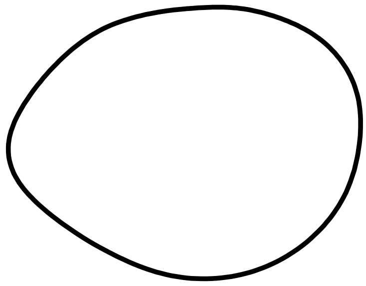
The rest of this course will concentrate on *algebraic* topology. The class of spaces we consider will often be more restrictive than up to this point, e.g. we will usually (though not always) require them to be Hausdorff, second countable, locally path-connected and one or two other conditions that are satisfied in all interesting examples.[6] It will happen often from now on that the best way to prove any given result is with a picture, but I might not always have time to produce the relevant picture in these notes. I'll do what I can.

As motivation, let us highlight two examples of questions that the tools of algebraic topology are designed to answer.

SAMPLE QUESTION 8.1. The following figures show two examples of **knots** K and K_0 in $\mathbb{R}^3$:



$K \subset \mathbb{R}^3$



$K_0 \subset \mathbb{R}^3$

The first knot K is known as the **trefoil** knot (*Kleeblattknoten*), and the second K_0 is the *trivial* knot or **unknot** (*Unknoten*). Roughly speaking, a knot is a subset in $\mathbb{R}^3$ that is homeomorphic to S^1 and satisfies some additional condition to avoid overly "wild" behavior, e.g. one could sensibly require each of K and K_0 to be the image of some infinitely differentiable 1-periodic map $\mathbb{R} \to \mathbb{R}^3$. The question then is: can K be deformed continuously to K_0? Let us express this more precisely. If you imagine K and K_0 as physical knots in space, then when you move them around, you don't

---

[6] The question of which examples are considered "interesting" depends highly on context, of course. In functional analysis, one encounters many interesting spaces of functions that do not have all of the properties we just listed. But this is not a course in functional analysis.

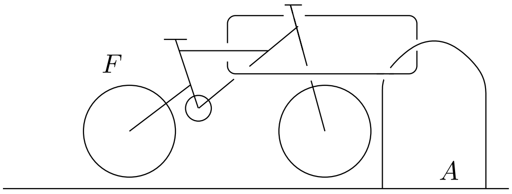
move only the knots—you also displace the air around them, and the motion of this collection of air particles over time can be viewed as defining a continuous family of homeomorphisms on $\mathbb{R}^3$. Mathematically, the question is then, does there exists a continuous map

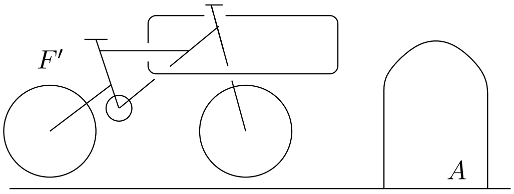
$$\varphi : [0,1] \times \mathbb{R}^3 \to \mathbb{R}^3$$

such that $\varphi(t,\cdot) : \mathbb{R}^3 \to \mathbb{R}^3$ is a homeomorphism for every $t \in [0,1]$, $\varphi(0,\cdot)$ is the identity map on $\mathbb{R}^3$ and $\varphi(1,\cdot) : \mathbb{R}^3 \to \mathbb{R}^3$ sends K_0 to K?

It turns out that the answer is no: in particular, if a homeomorphism $\varphi(1,\cdot)$ on $\mathbb{R}^3$ sending K_0 to K exists, then there must also be a homeomorphism between $\mathbb{R}^3 \backslash K$ and $\mathbb{R}^3 \backslash K_0$, and we will see that the latter is impossible. The reason is because we can associate to these spaces groups $\pi_1(\mathbb{R}^3 \backslash K)$ and $\pi_1(\mathbb{R}^3 \backslash K_0)$, which would need to be isomorphic if $\mathbb{R}^3 \backslash K$ and $\mathbb{R}^3 \backslash K_0$ were homeomorphic, and we will be able to compute enough information about both groups to show that they are not isomorphic.

Sample question 8.2. Here is another pair of spaces defined as subsets of $\mathbb{R}^3$:





A question of tremendous practical import: can the set F in the picture at the left be shifted continuously to match the set F' in the picture at the right, but without "passing through" A, i.e. is there a continuous family of embeddings $F \hookrightarrow \mathbb{R}^3 \backslash A$ that begins as the natural inclusion and ends by sending F to F'? If there is, then you may want to adjust your bike lock.

Of course there is no such continuous family of embeddings, and to see why, you could just delete the bicycle from the picture and pay attention only to the loop representing the bike lock, which is shown "linked" with A in the left picture and not in the right picture. The precise way to express the impossibility of deforming one picture to the other is that this loop is parametrized by a "noncontractible loop" $\gamma : S^1 \to \mathbb{R}^3 \backslash A$, meaning γ represents a nontrivial element in the fundamental group $\pi_1(\mathbb{R}^3 \backslash A)$.

Our task in this lecture is to define what the fundamental group is for an arbitrary space. We will then develop a few more of its general properties in the next lecture and spend the next four or five weeks developing methods to compute it.

We must first discuss paths in a space X. Since the unit interval $[0,1]$ will appear very often in the rest of this course, let us abbreviate it from now on by

$$I := [0,1].$$

For two points $x, y \in X$, a **path** (*Pfad*) from x to y is a map $\gamma : I \to X$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$.[7] We will sometimes use the notation

$$x \overset{\gamma}{\rightsquigarrow} y$$

to indicate that γ is a path from x to y.

The **inverse** of a path $x \overset{\gamma}{\rightsquigarrow} y$ is the path

$$y \overset{\gamma^{-1}}{\rightsquigarrow} x$$

---

[7] This seems a good moment to emphasize that all maps in this course are assumed continuous unless otherwise noted.

defined by $\gamma^{-1}(t) := \gamma(1-t)$. The reason for this terminology and notation will become clearer when we give the definition of the fundamental group below. The same goes for the notion of the **product** of two paths: there is no natural multiplication defined for a pair of paths between arbitrary points, but given $x \overset{\alpha}{\rightsquigarrow} y$ and $y \overset{\beta}{\rightsquigarrow} z$, we can define the product path $x \overset{\alpha\cdot\beta}{\rightsquigarrow} z$ by

$$(\alpha \cdot \beta)(t) = \begin{cases} \alpha(2t) & \text{if } 0 \leqslant t \leqslant 1/2, \\ \beta(2t-1) & \text{if } 1/2 \leqslant t \leqslant 1. \end{cases} \tag{8.1}$$

This operation is also called a **concatenation** of paths. The **trivial path** at a point $x \in X$ is defined as the constant path $x \overset{e_x}{\rightsquigarrow} x$, i.e.

$$e_x(t) = x.$$

The idea is for this to play the role of the identity element in some kind of group structure.

If we want to turn concatenation into a product structure on a group, then we have one immediate problem: it is not associative. In fact, given paths $x \overset{\alpha}{\rightsquigarrow} y$, $y \overset{\beta}{\rightsquigarrow} z$ and $z \overset{\gamma}{\rightsquigarrow} a$, we have

$$\alpha \cdot (\beta \cdot \gamma) \neq (\alpha \cdot \beta) \cdot \gamma,$$

though clearly the images of these two concatenations are the same, and their difference is only in the way they are parametrized. We would like to introduce an equivalence relation on the set of paths that forgets this distinction in parametrizations.

Definition 8.3. Two maps $f, g : X \to Y$ are **homotopic** (*homotop*) if there exists a map

$$H : I \times X \to Y \quad \text{such that } H(0, \cdot) = f \text{ and } H(1, \cdot) = g.$$

The map H is in this case called a **homotopy** (*Homotopie*) from f to g, and when a homotopy exists, we shall write

$$f \underset{h}{\sim} g.$$

It is straightforward to show that $\underset{h}{\sim}$ is an equivalence relation. In particular, if there are homotopies from f to g and from g to h, then by reparametrizing the parameter in $I = [0, 1]$ we can "glue" the two homotopies together to form a homotopy from f to h. The definition of the new homotopy is analogous to the definition of the concatenation of paths in (8.1).

For paths in particular we will need a slightly more restrictive notion of homotopy that fixes the end points.

Definition 8.4. For two paths α and β from x to y, we write

$$\alpha \underset{h+}{\sim} \beta$$

and say α is **homotopic with fixed end points** to β if there exists a map $H : I \times I \to X$ satisfying $H(0, \cdot) = \alpha$, $H(1, \cdot) = \beta$, $H(s, 0) = x$ and $H(s, 1) = y$ for all $s \in I$.

Exercise 8.5. Show that for any two points $x, y \in X$, $\underset{h+}{\sim}$ defines an equivalence relation on the set of all paths from x to y.

We will now prove several easy results about paths and homotopies. In most cases we will give precise formulas for the necessary homotopies, but one can also represent the main idea quite easily in pictures (see e.g. [**Hat02**, pp. 26–27]). We adopt the following convenient terminology: if $H : I \times X \to Y$ is a homotopy from $f_0 := H(0, \cdot) : X \to Y$ to $f_1 := H(1, \cdot) : X \to Y$, then we obtain a **continuous family of maps** $f_s := H(s, \cdot) : X \to Y$ for $s \in I$. The words "continuous family" will be understood as synonymous with "homotopy" in this sense.

PROPOSITION 8.6. *If $\alpha \underset{h+}{\sim} \alpha'$ are homotopic paths from x to y and $\beta \underset{h+}{\sim} \beta'$ are homotopic paths from y to z, then*

$$\alpha \cdot \beta \underset{h+}{\sim} \alpha' \cdot \beta'.$$

PROOF. By assumption, there exist continuous families of paths $x \overset{\alpha_s}{\rightsquigarrow} y$ and $y \overset{\beta_s}{\rightsquigarrow} z$ for $s \in I$ with $\alpha_0 = \alpha$, $\alpha_1 = \alpha'$, $\beta_0 = \beta$ and $\beta_1 = \beta'$. Then a homotopy with fixed end points from $\alpha \cdot \beta$ to $\alpha' \cdot \beta'$ can be defined via the continuous family

$$x \overset{\alpha_s \cdot \beta_s}{\rightsquigarrow} z \quad \text{for} \quad s \in I.$$

☐

We next show that while concatenation of paths is not an associative operation, it is associative "up to homotopy".

PROPOSITION 8.7. *Given paths $x \overset{\alpha}{\rightsquigarrow} y$, $y \overset{\beta}{\rightsquigarrow} z$ and $z \overset{\gamma}{\rightsquigarrow} a$,*

$$(\alpha \cdot \beta) \cdot \gamma \underset{h+}{\sim} \alpha \cdot (\beta \cdot \gamma).$$

PROOF. A suitable homotopy $H : I \times I \to X$ can be defined as a family of linear reparametrizations of the sequence of paths α, β, γ:

$$H(s,t) = \begin{cases} \alpha\left(\frac{4t}{s+1}\right) & \text{if } 0 \leqslant t \leqslant \frac{s+1}{4}, \\ \beta(4t - (s+1)) & \text{if } \frac{s+1}{4} \leqslant t \leqslant \frac{s+2}{4}, \\ \gamma\left(\frac{4}{2-s}(t-1) + 1\right) & \text{if } \frac{s+2}{4} \leqslant t \leqslant 1. \end{cases}$$

☐

And finally, a result that allows us to interpret the constant paths e_x as "identity elements" and γ and γ^{-1} as "inverses":

PROPOSITION 8.8. *For any path $x \overset{\gamma}{\rightsquigarrow} y$, the following relations hold:*

(i) $e_x \cdot \gamma \underset{h+}{\sim} \gamma$
(ii) $\gamma \underset{h+}{\sim} \gamma \cdot e_y$
(iii) $\gamma \cdot \gamma^{-1} \underset{h+}{\sim} e_x$
(iv) $\gamma^{-1} \cdot \gamma \underset{h+}{\sim} e_y$

PROOF. For (i), we define a family of reparametrizations of the concatenated path $e_x \cdot \gamma$ that shrinks the amount of time spent on e_x from 1/2 to 0:

$$H(s,t) = \begin{cases} x & \text{if } 0 \leqslant t \leqslant \frac{1-s}{2}, \\ \gamma\left(\frac{2}{s+1}(t-1) + 1\right) & \text{if } \frac{1-s}{2} \leqslant t \leqslant 1. \end{cases}$$

The homotopy for (ii) is analogous.

For (iii), the idea is to define a family of paths that traverse only part of γ up to some time depending on s, then stay still for a suitable length of time and, in a third step, follow γ^{-1} back to x:

$$H(s,t) = \begin{cases} \gamma(2t) & \text{if } 0 \leqslant t \leqslant \frac{1-s}{2}, \\ \gamma(1-s) & \text{if } \frac{1-s}{2} \leqslant t \leqslant \frac{1+s}{2}, \\ \gamma(2-2t) & \text{if } \frac{1+s}{2} \leqslant t \leqslant 1. \end{cases}$$

The last relation follows from this by interchanging the roles of γ and γ^{-1}. ☐

The last three propositions combine to imply that the group structure in the following definition is a well-defined associative product which admits an identity element and inverses.

Definition 8.9. Given a space X and a point $p \in X$, the **fundamental group** (*Fundamentalgruppe*) of X with **base point** (*Basispunkt*) p is defined as the set of equivalence classes of paths $p \rightsquigarrow p$ up to homotopy with fixed end points:

$$\pi_1(X, p) := \left\{ \text{paths } p \overset{\gamma}{\rightsquigarrow} p \right\} \Big/ \underset{h+}{\sim} .$$

The product of two equivalence classes $[\alpha], [\beta] \in \pi_1(X, p)$ is defined via concatenation:

$$[\alpha][\beta] := [\alpha \cdot \beta],$$

and the identity element is represented by the constant path $[e_p]$. The inverse element for $[\gamma] \in \pi_1(X, p)$ is represented by the reversed path γ^{-1}.

Before exploring the further properties of the group $\pi_1(X, p)$, let us clarify in what sense it is a "topological invariant" of the space X. Intuitively, we would like this to mean that whenever X and Y are two homeomorphic spaces, their fundamental groups should be isomorphic groups. What makes this statement a tiny bit more complicated is that the fundamental group of X doesn't just depend on X alone, but also on a choice of base point, so in order to make precise and correct statements about topological invariance, we will need to carry around a base point as extra data. The following definition is intended to formalize this notion.

Definition 8.10. A **pointed space** (*punktierter Raum*) is a pair (X, p) consisting of a topological space X and a point $p \in X$. The point $p \in X$ is in this case called the **base point** (*Basispunkt*) of X. Given pointed spaces (X, p) and (Y, q), any continuous map $f : X \to Y$ satisfying $f(p) = q$ is called a **pointed map** or **map of pointed spaces**, and can be denoted by

$$f : (X, p) \to (Y, q).$$

We also sometimes refer to such objects as **base-point preserving** maps. Finally, given two pointed maps $f, g : (X, p) \to (Y, q)$, a homotopy $H : I \times X \to Y$ from f to g that satisfies $H(s, p) = q$ for all $s \in I$ is called a **pointed homotopy**, or **homotopy of pointed maps**, or **base-point preserving homotopy**. One can equivalently describe such a homotopy as a continuous 1-parameter family of pointed maps $f_s := H(s, \cdot) : (X, p) \to (Y, q)$ defined for $s \in I$.

Here is the first main result about the topological invariance of π_1:

Theorem 8.11. *One can associate to every pointed map $f : (X, p) \to (Y, q)$ a group homomorphism*

$$f_* : \pi_1(X, p) \to \pi_1(Y, q) : [\gamma] \mapsto [f \circ \gamma],$$

which has the following properties:

(i) For any pointed maps $(X, p) \xrightarrow{f} (Y, q)$ and $(Y, q) \xrightarrow{g} (Z, r)$, $(g \circ f)_* = g_* \circ f_*$.

(ii) The map associated to the identity map $(X, p) \xrightarrow{\mathrm{Id}} (X, p)$ is the identity homomorphism $\pi_1(X, p) \xrightarrow{\mathbb{1}} \pi_1(X, p)$.

(iii) Each homomorphism f_* depends only on the pointed homotopy class of f.

Proof. It is clear that up to homotopy (with fixed end points), the path $q \overset{f \circ \gamma}{\rightsquigarrow} q$ in Y depends only on the path $p \overset{\gamma}{\rightsquigarrow} p$ only up to homotopy with fixed end points; indeed, if $H : I \times I \to X$ defines a homotopy with fixed end points between two paths α and β based at p, then $f \circ H : I \times I \to Y$ defines a corresponding homotopy between $f \circ \alpha$ and $f \circ \beta$. Similarly, if $[\gamma] \in \pi_1(X, p)$ and $f, g : (X, p) \to (Y, q)$ are homotopic via a base-point preserving homotopy $H : I \times X \to Y$, then

$h : I \times I \to Y : (s,t) \mapsto H(s, \gamma(t))$ defines a homotopy with fixed end points between $f \circ \gamma$ and $g \circ \gamma$. This shows that $f_* : \pi_1(X,p) \to \pi_1(Y,q)$ is a well-defined map that depends on f only up to base-point preserving homotopy. It is similarly easy to check that f_* is a homomorphism and satisfies the first two stated properties: e.g. for any two paths $p \overset{\alpha,\beta}{\rightsquigarrow} p$, we have

$$f_*([\alpha][\beta]) = [f \circ (\alpha \cdot \beta)] = [(f \circ \alpha) \cdot (f \circ \beta)] = f_*[\alpha] f_*[\beta]$$

and

$$f_*[e_p] = [e_q].$$

□

Corollary 8.12. *If X and Y are spaces admitting a homeomorphism $f : X \to Y$, then for any choice of base point $p \in X$, the groups $\pi_1(X,p)$ and $\pi_1(Y, f(p))$ are isomorphic.*

Proof. Abbreviate $q := f(p)$, so $f : (X,p) \to (Y,q)$ is a pointed map, and since its inverse is continuous, $f^{-1} : (Y,q) \to (X,p)$ is also a pointed map. Using Theorem 8.11, the commutative diagram (see Remark 8.14 below) of continuous maps

(8.2)
$$\begin{tikzcd}
& (Y,q) \arrow[dr, "f^{-1}"] & \\
(X,p) \arrow[ur, "f"] \arrow[rr, "\mathrm{Id}"] & & (X,p)
\end{tikzcd}$$

then gives rise to a similar commutative diagram of group homomorphisms

(8.3)
$$\begin{tikzcd}
& \pi_1(Y,q) \arrow[dr, "f_*^{-1}"] & \\
\pi_1(X,p) \arrow[ur, "f_*"] \arrow[rr, "\mathbb{1}"] & & \pi_1(X,p)
\end{tikzcd}$$

Reversing the roles of (X,p) and (Y,q) produces similar diagrams to show that f_* and f_*^{-1} are inverse homomorphisms, hence both are isomorphisms. □

Remark 8.13. The fancy way to summarize Theorem 8.11 is that π_1 defines a "covariant functor" from the category of pointed spaces and pointed homotopy classes to the category of groups and homomorphisms. We will discuss categories and functors more next semester in *Topologie II*.

Remark 8.14. Commutative diagrams such as (8.2) and (8.3) will appear more and more often as we get deeper into algebraic topology. When we say that such a diagram **commutes**, it means that any two maps obtained by composing a sequence of arrows along different paths from one place in the diagram to another must match, so e.g. the message carried by (8.2) is the relation $f^{-1} \circ f = \mathrm{Id}$, and (8.3) means $f_*^{-1} \circ f_* = \mathbb{1}$. These were especially simple examples, but later we will also encounter larger diagrams like

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C_* \arrow[d, "\gamma"] \\
A \arrow[r, "f'"] & B' \arrow[r, "g'"] & C'
\end{tikzcd}$$

The purpose of this one is to communicate the two relations $\beta \circ f = f' \circ \alpha$ and $\gamma \circ g = g' \circ \beta$, along with all the more complicated relations that follow from these, such as $g' \circ f' \circ \alpha = \gamma \circ g \circ f$.

Since the paths representing elements of $\pi_1(X, p)$ have the same fixed starting and ending point, we often think of them as *loops* in X. We will establish some general properties of $\pi_1(X, p)$ in the next lecture, starting with the observation that whenever X is path-connected, $\pi_1(X, p)$ up to isomorphism does not actually depend on the choice of the base point $p \in X$, thus we can sensibly write it as $\pi_1(X)$. Computing $\pi_1(X)$ for a given space X is not always easy or possible, but we will develop some methods that are very effective on a wide class of spaces. I can already mention two simple examples: first, $\pi_1(\mathbb{R}^n)$ is the trivial group, resulting from the relatively obvious fact that (by linear interpolation) every path in $\mathbb{R}^n$ from a point to itself is homotopic with fixed end points to the constant path. In contrast, we will see that $\pi_1(S^1)$ and $\pi_1(\mathbb{R}^2 \backslash \{0\})$ are both isomorphic to the integers, and this simple result already has many useful applications, e.g. we will derive from it a very easy proof of the fundamental theorem of algebra.

9. Some properties of the fundamental group

We would now like to clarify to what extent $\pi_1(X, p)$ depends on p in addition to X.

Theorem 9.1. *Given $p, q \in X$, any homotopy class (with fixed end points) of paths $p \overset{\gamma}{\rightsquigarrow} q$ determines a group isomorphism*

$$\Phi_\gamma : \pi_1(X, q) \to \pi_1(X, p) : [\alpha] \mapsto [\gamma \cdot \alpha \cdot \gamma^{-1}].$$

Proof. Note that in writing the formula above for $\Phi_\gamma([\alpha])$, we are implicitly using the fact (Proposition 8.7) that concatenation of paths is an associative operation up to homotopy, so one can represent $\Phi_\gamma([\alpha])$ by either of the paths $\gamma \cdot (\alpha \cdot \gamma^{-1})$ or $(\gamma \cdot \alpha) \cdot \gamma^{-1}$ without the result depending on this choice. Similarly, Proposition 8.6 implies that the homotopy class of $\gamma \cdot \alpha \cdot \gamma^{-1}$ with fixed end points only depends on the homotopy classes of γ and α (also with fixed end points).[8] This proves that Φ_γ is a well-defined map as written. The propositions in the previous lecture imply in a similarly straightforward manner that Φ_γ is a homomorphism, i.e.

$$\Phi_\gamma([\alpha][\beta]) = [\gamma \cdot \alpha \cdot \beta \cdot \gamma^{-1}] = [\gamma \cdot \alpha \cdot \gamma^{-1} \cdot \gamma \cdot \beta \cdot \gamma^{-1}] = \Phi_\gamma([\alpha])\Phi_\gamma([\beta]),$$

and

$$\Phi_\gamma([e_q]) = [\gamma \cdot e_q \cdot \gamma^{-1}] = [\gamma \cdot \gamma^{-1}] = [e_p].$$

It remains only to observe that Φ_γ and $\Phi_{\gamma^{-1}}$ are inverses of each other, hence both are isomorphisms. ☐

Corollary 9.2. *If X is path-connected, then $\pi_1(X, p)$ up to isomorphism is independent of the choice of base point $p \in X$.* ☐

Due to this corollary, it is conventional to abbreviate the fundamental group by

$$\pi_1(X) := \pi_1(X, p)$$

whenever X is path-connected, and we will see many theorems about $\pi_1(X)$ in situations where the base point plays no important role. If X is not path-connected but $X_0 \subset X$ denotes the path-component containing p, then $\pi_1(X, p) = \pi_1(X_0, p) \cong \pi_1(X_0)$, so in practice it is sufficient to restrict our attention to path-connected spaces. Some caution is nonetheless warranted in using the notation $\pi_1(X)$: strictly speaking, $\pi_1(X)$ is not a concrete group but only an isomorphism class of groups, and the subtle distinction between these two notions occasionally leads to trouble. You should always keep in the back of your mind that even if the base point is not mentioned, it is an essential piece of the definition of $\pi_1(X)$.

---

[8] Note that the homotopy class of γ determines that of γ^{-1}. (Why?)

We next discuss some alternative ways to interpret $\pi_1(X,p)$. Recall the following useful notational device: given a space X with subset $A \subset X$, we define

$$X/A := X/\sim$$

with the quotient topology, where the equivalence relation defines $a \sim b$ for all $a, b \in A$. In other words, this is the quotient space obtained from X by "collapsing" the subset A to a single point. For example, it is straightforward (see Exercise 5.16) to show that $\mathbb{D}^n/S^{n-1}$ is homeomorphic to S^n for every $n \in \mathbb{N}$, and if we replace $\mathbb{D}^1 = [-1,1]$ by the unit interval $I = [0,1]$, we obtain the special case

$$[0,1]/\{0,1\} = I/\partial I \cong S^1.$$

Here we have used the notation

$$\partial X := \text{"boundary of } X\text{"},$$

which comes from differential geometry, so for instance $\partial\mathbb{D}^n = S^{n-1}$ and we can therefore also identify S^n with $\mathbb{D}^n/\partial\mathbb{D}^n$. A specific homeomorphism $I/\partial I \to S^1$ can be written most easily by thinking of S^1 as the unit circle in $\mathbb{C}$:

$$I/\partial I \to S^1 : [t] \mapsto e^{2\pi i t}.$$

LEMMA 9.3. *For any space X and subset $A \subset X$, there is a canonical bijection between the set of all continuous maps $f : X \to Y$ that are constant on A and the set of all continuous maps $g : X/A \to Y$. For any two maps f and g that correspond under this bijection, the diagram*

$$\begin{tikzcd} X \arrow[rr, "\pi"] \arrow[dr, "f"'] & & X/A \arrow[dl, "g"] \\ & Y & \end{tikzcd}$$

commutes, where $\pi : X \to X/A$ denotes the quotient projection; in other words, $g \circ \pi = f$.

PROOF. The diagram determines the correspondence: given $g : X/A \to Y$, we can define $f := g \circ \pi$ to obtain a map $X \to Y$ that is automatically constant on A, and conversely, if $f : X \to Y$ is given and is constant on A, then there is a well-defined map $g : X/A \to Y : [x] \mapsto f(x)$. Our main task is to show that f is continuous if and only if g is continuous. In one direction this is immediate: if g is continuous, then $f = g \circ \pi$ is the composition of two continuous maps and is therefore also continuous. Conversely, if f is continuous, then for every open set $\mathcal{U} \subset Y$, we know $f^{-1}(\mathcal{U}) \subset X$ is open. A point $[x] \in X/A$ is then in $g^{-1}(\mathcal{U})$ if and only if $x \in f^{-1}(\mathcal{U})$, so $g^{-1}(\mathcal{U}) = \pi(f^{-1}(\mathcal{U}))$ and thus $\pi^{-1}(g^{-1}(\mathcal{U})) = f^{-1}(\mathcal{U})$ is open. By the definition of the quotient topology, this means that $g^{-1}(\mathcal{U}) \subset X/A$ is open, so g is continous. $\square$

Lemma 9.3 gives a canonical bijection between the set of all paths $p \overset{\gamma}{\rightsquigarrow} p$ in X beginning and ending at the base point and the set of all continuous pointed maps

$$(I/\partial I, [0]) \to (X, p).$$

It is easy to check moreover that two paths $p \overset{\gamma}{\rightsquigarrow} p$ are homotopic with fixed end points if and only if they correspond to maps $(I/\partial I, [0]) \to (X,p)$ in the same *pointed* homotopy class. Under the aforementioned homeomorphism $I/\partial I \cong S^1 \subset \mathbb{C}$ that identifies $[0] = [1]$ with 1, this gives us an alternative description of $\pi_1(X,p)$ as

$$\pi_1(X,p) = \{\text{pointed maps } \gamma : (S^1, 1) \to (X,p)\} \Big/ \underset{h+}{\sim},$$

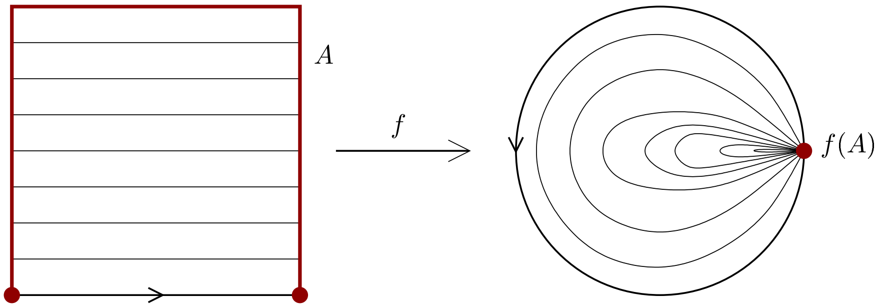


FIGURE 1. A map $f : I^2 \to \mathbb{D}^2$ which descends to a homeomorphism $g : I^2/A \to \mathbb{D}^2$ in the proof of Theorem 9.4.

where $\underset{h+}{\sim}$ now denotes the equivalence relation defined by pointed homotopy. The group structure of $\pi_1(X, p)$ is less easy to see from this perspective, but it will nonetheless be extremely useful to think of elements of $\pi_1(X)$ as represented by *loops* $\gamma : S^1 \to X$.

THEOREM 9.4. *A loop $\gamma : (S^1, 1) \to (X, p)$ represents the identity element in $\pi_1(X, p)$ if and only if there exists a continuous map $u : \mathbb{D}^2 \to X$ with $u|_{\partial\mathbb{D}^2} = \gamma$.*

PROOF. I can't explain this proof without a picture, so to start with, have a look at Figure 1. It depicts a map $f : I^2 \to \mathbb{D}^2 \subset \mathbb{C}$ that collapses the red region consisting of three sides of the square

$$A := (\partial I \times I) \cup (I \times \{1\}) \subset I^2$$

to the single point $f(A) = \{1\} \subset \mathbb{D}^2$, but is bijective everywhere else, and maps the path $I \times \{0\} \subset I^2$ to the loop $\partial\mathbb{D}^2$. By Lemma 9.3, f determines a map

$$g : I^2/A \to \mathbb{D}^2$$

which is continuous and bijective, and it is also an open map (i.e. it maps open sets to open sets), hence its inverse is also continuous and g is therefore a homeomorphism. Now, a path $\gamma : I \to X$ with $\gamma(0) = \gamma(1) = p$ represents the identity in $\pi_1(X, p)$ if and only if there exists a homotopy $H : I^2 \to X$ with $H(0, \cdot) = \gamma$ and $H|_A \equiv p$. Applying Lemma 9.3 again, such a map is equivalent to a map $h : I^2/A \to X$ which sends the equivalence class represented by every point in A to the base point p. In this case, $h \circ g^{-1}$ is a map $\mathbb{D}^2 \to X$ whose restriction to $\partial\mathbb{D}^2$ is the loop $S^1 \cong I/\partial I \to X$ determined by $\gamma : I \to X$. ☐

REMARK 9.5. Maps $\gamma : S^1 \to X$ that admit extensions over $\mathbb{D}^2$ as in the above theorem are called **contractible loops** (*zusammenziehbare Schleifen*).

DEFINITION 9.6. A space X is called **simply connected** (*einfach zusammenhängend*) if it is path-connected and its fundamental group is trivial.

It is common to denote the trivial group by "0", so for path-connected spaces, we can write

$$X \text{ is simply connected} \quad \Leftrightarrow \quad \pi_1(X) = 0.$$

By Theorem 9.4, this is equivalent to the condition that every map $\gamma : S^1 \to X$ admits a continuous extension $u : \mathbb{D}^2 \to X$ satisfying $u|_{\partial\mathbb{D}^2} = \gamma$. Note that there was no need to mention the base point in this formulation: if X is path-connected, then $\pi_1(X) = 0$ means $\pi_1(X, p) = 0$ for every p, so for a given loop $\gamma : S^1 \to X$ we are free to choose $p := \gamma(1) \in X$ as the base point and then apply Theorem 9.4.

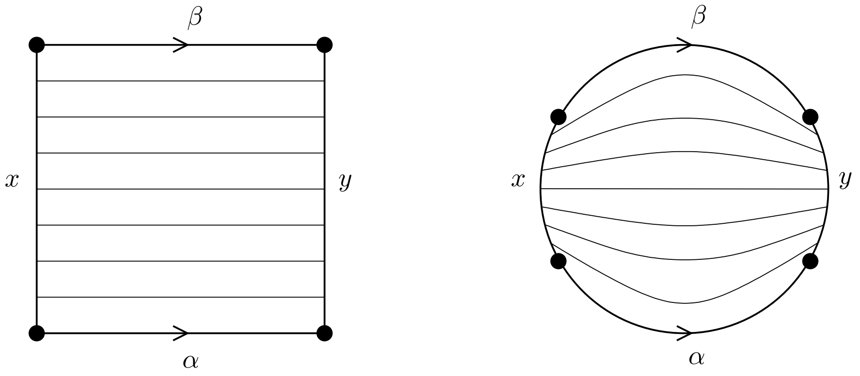


Figure 2. Two equivalent pictures of the same homotopy with fixed end points x and y between two paths α and β, using a homeomorphism $I^2 \cong \mathbb{D}^2$.

Examples 9.7. Though we will need to develop a few more tools before we can prove it, the sphere S^2 is simply connected. (Try to imagine a loop in S^2 that cannot be filled in by a disk—but do not try too hard!)

In contrast, $\mathbb{R}^2 \backslash \{0\}$ is not simply connected: we will see that the natural inclusion map $\gamma : S^1 \hookrightarrow \mathbb{R}^2 \backslash \{0\}$ is an example of a loop that cannot be extended to a map $u : \mathbb{D}^2 \to \mathbb{R}^2 \backslash \{0\}$. Of course, it *can* be extended to a map $\mathbb{D}^2 \to \mathbb{R}^2$, but it will turn out that such an extension must always hit the origin somewhere—in other words, the loop is contractible in $\mathbb{R}^2$, but not contractible in $\mathbb{R}^2 \backslash \{0\}$. This observation has many powerful implications, e.g. we will see in the next lecture that it is the key idea behind one of the simplest proofs of the *fundamental theorem of algebra*, that every nonconstant complex polynomial has a root.

Another example with nontrivial fundamental group is the **torus** $\mathbb{T}^2 := S^1 \times S^1$. Pictures of this space embedded in $\mathbb{R}^3$ typically depict it as the surface of a tube (or a doughnut or a bagel—depending on your cultural preferences). Can you visualize a loop on this surface that is contractible in $\mathbb{R}^3$ but not in $\mathbb{T}^2$?

One can also use the fundamental group to gain insight into homotopy classes of non-closed paths:

Theorem 9.8. *Two paths* $x \overset{\alpha,\beta}{\rightsquigarrow} y$ *in* X *are homotopic with fixed end points if and only if the concatenated path* $x \overset{\alpha\cdot\beta^{-1}}{\rightsquigarrow} x$ *represents the identity element in* $\pi_1(X, x)$.

Proof. The condition $\alpha \underset{h+}{\sim} \beta$ means the existence of a homotopy $H : I^2 \to X$ with certain properties as depicted at the left in Figure 2, but by a suitable choice of homeomorphism $I^2 \cong \mathbb{D}^2$ as shown to the right of that picture, we can equally well regard H as a map $\mathbb{D}^2 \to X$. The loop $\gamma := H|_{\partial\mathbb{D}^2} : S^1 \to X$ can then be viewed as the concatenation $\alpha \cdot e_y \cdot \beta^{-1} \cdot e_x$, which by Proposition 8.8 is homotopic with fixed end points to $\alpha \cdot \beta^{-1}$. The result then follows directly from Theorem 9.4. ☐

Corollary 9.9. *A space* X *is simply connected if and only if for every pair of points* $p, q \in X$*, there exists a path from* p *to* q *and it is unique up to homotopy with fixed end points.* ☐

Let us finally work out a few concrete examples.

Example 9.10. For each $n \geqslant 0$, the Euclidean space $\mathbb{R}^n$ is simply connected. Indeed, since it is path-connected, we are free to choose the base point $0 \in \mathbb{R}^n$, and can then observe that every loop $0 \overset{\gamma}{\rightsquigarrow} 0$ is homotopic to the constant loop via the continuous family of loops

$$\gamma_s : I \to \mathbb{R}^n : t \mapsto s\gamma(t) \quad \text{for} \quad s \in I.$$

Example 9.11. Since every open ball $B_r(x)$ in $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^n$ itself, Corollary 8.12 implies that $\pi_1(B_r(x))$ also vanishes, i.e. $B_r(x)$ is simply connected. One could also give a direct proof of this, analogously to Example 9.10: just choose $x \in B_r(x)$ as the base point and define γ_s via linear interpolation between γ and the constant loop at x. A similar trick works in fact for any *convex* subset $K \subset \mathbb{R}^n$, i.e. any set K with the property that the straight line segment connecting any two points $x, y \in K$ is also contained in K. It follows that all convex subsets of finite-dimensional vector spaces are simply connected.

Example 9.12. Our first example of a nontrivial fundamental group (and probably also the most important one to take note of in this course) is the circle: we claim that

$$\pi_1(S^1) \cong \mathbb{Z}.$$

The proof is based on a pair of lemmas that we will prove (in more general forms) in a few weeks, though I suspect you will already find them easy to believe. Regarding S^1 as the unit circle in $\mathbb{C}$, consider the map

$$f : \mathbb{R} \to S^1 : t \mapsto e^{2\pi i t}.$$

This is our first interesting example of a so-called **covering map** (*Überlagerung*): it is surjective, and it looks like a homeomorphism *on the small scale* (i.e. if you zoom in close enough on any particular point in $\mathbb{R}$), but it is not injective, in fact it "wraps" the line $\mathbb{R}$ around S^1 infinitely many times. The next two statements are special cases of results that we will later prove about a much more general class of covering spaces:

(1) Given a path $x \overset{\gamma}{\rightsquigarrow} y$ in S^1 and a point $\tilde{x} \in f^{-1}(x)$, there exists a unique path $\tilde{x} \overset{\tilde{\gamma}}{\rightsquigarrow} \tilde{y}$ in $\mathbb{R}$ that is a "lift" of γ in the sense that $f \circ \tilde{\gamma} = \gamma$.
(2) Given a homotopy $H : I \times I \to S^1$ of paths $x \overset{\gamma}{\rightsquigarrow} y$ (with fixed end points) and a point $\tilde{x} \in f^{-1}(x)$, there exists a unique homotopy $\widetilde{H} : I \times I \to \mathbb{R}$ of lifted paths $\tilde{x} \overset{\tilde{\gamma}}{\rightsquigarrow} \tilde{y}$ which lifts H in the sense that $f \circ \widetilde{H} = H$.

Now for any $[\gamma] \in \pi_1(S^1, 1)$ represented by a path $1 \overset{\gamma}{\rightsquigarrow} 1$, there is a unique lift to a path $0 \overset{\tilde{\gamma}}{\rightsquigarrow} \tilde{\gamma}(1)$ in $\mathbb{R}$. Unlike γ, the end point of the lift need not match its starting point, but the fact that it is a lift implies $\tilde{\gamma}(1) \in f^{-1}(1) = \mathbb{Z}$, and the fact that homotopies can be lifted implies that this integer does not change if we replace γ with any other representative of $[\gamma] \in \pi_1(S^1, 1)$. We therefore obtain a well-defined map

$$\Phi : \pi_1(S^1, 1) \to \mathbb{Z} : [\gamma] \mapsto \tilde{\gamma}(1).$$

It is easy to show that Φ is a group homomorphism by lifting concatenated paths. Moreover, Φ is surjective since $\Phi([\gamma_k]) = k$ for each of the loops $\gamma_k(t) = e^{2\pi i k t}$ with $k \in \mathbb{Z}$, as these have lifts $\tilde{\gamma}(t) = kt$. Injectivity amounts to the statement that γ must be homotopic to a constant whenever its lift satisfies $\tilde{\gamma}(1) = 0$, and this follows from the fact that $\pi_1(\mathbb{R}) = 0$: indeed, in this case $\tilde{\gamma}$ is not just a path in $\mathbb{R}$ but is also a loop, thus it represents an element of $\pi_1(\mathbb{R}, 0) = 0$ and is therefore homotopic to the constant loop. Composing that homotopy with $f : \mathbb{R} \to S^1$ gives a homotopy of the original loop γ to a constant.

Exercise 9.13. In this exercise we show that the fundamental group of a product is a product of fundamental groups.

(a) Given two pointed spaces (X, x) and (Y, y), prove that $\pi_1(X \times Y, (x, y))$ is isomorphic to the product group $\pi_1(X, x) \times \pi_1(Y, y)$.

Hint: Use the projections $p^X : X \times Y \to X$ *and* $p^Y : X \times Y \to Y$ *to define a natural map from* π_1 *of the product to the product of* π_1*'s, then prove that it is an isomorphism.*

(b) Generalize part (a) to the case of an infinite product of pointed spaces (with the product topology).

Exercise 9.14. Let us regard $\pi_1(X, p)$ as the set of base-point preserving homotopy classes of maps $(S^1, \mathrm{pt}) \to (X, p)$, and let $[S^1, X]$ denote the set of homotopy classes of maps $S^1 \to X$, with no conditions on base points. (The elements of $[S^1, X]$ are called **free homotopy classes** of loops in X). There is a natural map

$$F : \pi_1(X, p) \to [S^1, X]$$

defined by ignoring base points. Prove:

(a) F is surjective if X is path-connected.

(b) $F([\alpha]) = F([\beta])$ if and only if $[\alpha]$ and $[\beta]$ are conjugate in $\pi_1(X, p)$.
Hint: If $H : [0, 1] \times S^1 \to X$ *is a homotopy with* $H(0, \cdot) = \alpha$ *and* $H(1, \cdot) = \beta$, *and* $t_0 \in S^1$ *is the base point in* S^1, *then* $\gamma := H(\cdot, t_0) : [0, 1] \to X$ *begins and ends at* p, *and therefore also defines a loop. Compare* α *and the concatenation* $\gamma \cdot \beta \cdot \gamma^{-1}$.

The conclusion is that if X is path-connected, F induces a bijection between $[S^1, X]$ and the set of conjugacy classes in $\pi_1(X)$. In particular, $\pi_1(X) \cong [S^1, X]$ whenever $\pi_1(X)$ is abelian.

10. Retractions and homotopy equivalence

Having proved that two homeomorphic spaces always have isomorphic fundamental groups, it is natural to wonder whether the converse is true. The answer is an emphatic *no*, but this will turn out to be more of an advantage than a disadvantage: it becomes much easier to compute $\pi_1(X)$ if we are free to replace X with another space X' that is not homeomorphic to X but still has certain features in common. This idea leads us naturally to the notion of *homotopy equivalence*, another equivalence relation on topological spaces that is strictly weaker than homeomorphism.

Let us first discuss conditions that make the homomorphisms $f_* : \pi_1(X, p) \to \pi_1(Y, q)$ injective or surjective.

Definition 10.1. For a space X with subset $A \subset X$, a map $f : X \to A$ is called a **retraction** (*Retraktion*) if $f|_A$ is the identity map $A \to A$. Equivalently, if $i : A \hookrightarrow X$ denotes the natural inclusion map, then f being a retraction means that the following diagram commutes:

(10.1)
$$\begin{tikzcd} A \arrow[rr, "\mathrm{Id}"] \arrow[dr, hook, "i"'] & & A \\ & X \arrow[ur, "f"'] & \end{tikzcd}$$

We say in this case that A is a **retract** of X.

Example 10.2. For $A := \mathbb{R} \times \{0\} \subset \mathbb{R}^2$, the map $f : \mathbb{R}^2 \to A : (x, y) \mapsto (x, 0)$ is a retraction.

A wide class of examples of retractions arises from the following general construction.

Definition 10.3. The **wedge sum** of two pointed spaces (X, p) and (Y, q) is the space

$$X \vee Y := (X \amalg Y)/\sim$$

where the equivalence relation sets $p \in X$ equivalent to $q \in Y$ and is otherwise trivial. More generally, any (potentially infinite) collection of pointed spaces $\{(X_\alpha, p_\alpha)\}_{\alpha \in J}$ has a wedge sum

$$\bigvee_{\alpha \in J} X_\alpha := \coprod_{\alpha \in J} X_\alpha \Big/ \sim,$$

where the equivalence relation identifies all the base points $p_\alpha \sim p_\beta$ for $\alpha, \beta \in J$. The wedge sum is naturally also a pointed space, with base point $[p_\alpha] \in \bigvee_\beta X_\beta$.

Remark 10.4. I did not specify the topology on $X \vee Y$ or $\bigvee_\alpha X_\alpha$, but by now you know enough to deduce from context what it must be: e.g. for the wedge of two spaces, we assign the disjoint union topology to $X \amalg Y$ and then endow $(X \amalg Y)/\sim$ with the resulting quotient topology. We will see many more constructions of this sort that involve a combination of quotients with disjoint unions and/or products, so you should always assume unless otherwise specified that the topology is whatever arises naturally from disjoint union, product and/or quotient topologies.

The notation for wedge sums is slightly nonideal since the definition of $\bigvee_\alpha X_\alpha$ depends not just on the spaces X_α but also on their base points $p_\alpha \in X_\alpha$, and it is not true in general that changing base points always produces homeomorphic wedge sums. It is true however for most examples that arise in practice, so the ambiguity in notation will usually not cause a problem. Note that since each of the individual spaces X_α are naturally subspaces of $\coprod_\beta X_\beta$, they can equally well be regarded as subspaces of $\bigvee_\beta X_\beta$, and it is straightforward to show that the obvious inclusion $X_\alpha \hookrightarrow \bigvee_\beta X_\beta$ for each α is a homeomorphism onto its image. As subspaces of a disjoint union $\coprod_\alpha X_\alpha$, the individual spaces X_β and X_γ for $\beta \neq \gamma$ are by definition disjoint, whereas in $\bigvee_\alpha X_\alpha$, they intersect each other at the base point, and only there.

Exercise 10.5. Show that for any collection of pointed maps $\{f_\alpha : (X_\alpha, p_\alpha) \to (Y, q)\}_{\alpha \in J}$, the unique map $f : \bigvee_{\alpha \in J} X_\alpha \to Y$ determined by the condition $f|_{X_\alpha} = f_\alpha$ for each $\alpha \in J$ is continuous.

Example 10.6. For the wedge sum $X \vee Y$ of two pointed spaces (X, p) and (Y, q), there is a natural base-point preserving retraction

$$f : X \vee Y \to X : [x] \mapsto \begin{cases} x & \text{if } x \in X, \\ p & \text{if } x \in Y. \end{cases}$$

In words, f maps $X \subset X \vee Y$ to itself as the identity map while collapsing all of $Y \subset X \vee Y$ to the base point. One can analogously define a natural retraction $X \vee Y \to Y$, and for a wedge sum of arbitrarily many spaces, a natural retraction $\bigvee_{\beta \in J} X_\beta \to X_\alpha$ for each $\alpha \in J$.

Exercise 10.7. Convince yourself that the map $f : X \vee Y \to X$ in Example 10.6 is continuous.

Example 10.8. For $X = Y = S^1$, the wedge sum $S^1 \vee S^1$ is a space homeomorphic to the symbols "8" and "∞", i.e. a so-called *figure eight*. Note that in this case, we did not need to specify the base points on the two copies of S^1 because choosing different base points leads to wedge sums that are homeomorphic. As a special case of Example 10.6, there are two retractions $S^1 \vee S^1 \to S^1$ that collapse either the top half or the bottom half of the "8" to a point.

The next example originates in the proof of the Brouwer fixed point theorem that we sketched at the end of Lecture 1 (cf. Theorem 1.13).

Example 10.9. As explained in Lecture 1, if there exists a continuous map $f : \mathbb{D}^n \to \mathbb{D}^n$ with no fixed point, then one can use it to define a map $g : \mathbb{D}^n \to \partial\mathbb{D}^n = S^{n-1}$ that satisfies $g(x) = x$ for all $x \in \partial\mathbb{D}^n$. The idea is to follow the unique line from x through $f(x)$ until arriving at some point of the boundary, which is defined to be $g(x)$. This makes g a retraction of $\mathbb{D}^n$ to $\partial\mathbb{D}^n$. The main step in the proof of Brouwer's fixed point theorem is to show that no such retraction exists. We will carry this out for $n = 2$ in a moment.

Theorem 10.10. *If $f : X \to A$ is a retraction and $i : A \hookrightarrow X$ denotes the inclusion, then for any choice of base point $a \in A$, the induced homomorphism $i_* : \pi_1(A, a) \to \pi_1(X, a)$ is injective, while $f_* : \pi_1(X, a) \to \pi_1(A, a)$ is surjective.*

Proof. Since the maps in the commutative diagram (10.1) all send the base point $a \in A$ to itself, Theorem 8.11 produces a corresponding commutative diagram of homomorphisms:

$$\begin{tikzcd}
\pi_1(A,a) \arrow[rr, "\mathbb{1}"] \arrow[dr, "i_*"] & & \pi_1(A,a) \\
& \pi_1(X,a) \arrow[ur, "f_*"] &
\end{tikzcd}$$

In particular, $f_* \circ i_*$ is both injective and surjective, which is only possible if i_* is injective and f_* is surjective. $\square$

Proof of the Brouwer fixed point theorem for $n = 2$. If there is a map $f : \mathbb{D}^2 \to \mathbb{D}^2$ with no fixed point, then there is also a retraction $g : \mathbb{D}^2 \to \partial\mathbb{D}^2 = S^1$ as explained in Example 10.9, so Theorem 10.10 implies that the induced homomorphism $g_* : \pi_1(\mathbb{D}^2) \to \pi_1(S^1)$ is surjective. As we saw at the end of the previous lecture, $\pi_1(S^1) \cong \mathbb{Z}$, and an easy modification of Example 9.10 shows that $\pi_1(\mathbb{D}^2) = 0$. (In fact, the same argument proves that every convex subset of $\mathbb{R}^n$ is simply connected—this will also follow from the more general Corollary 10.24 below.) But there is no surjective homomorphism from the trivial group to $\mathbb{Z}$, so this is a contradiction. $\square$

Definition 10.11. Assume X is a space with subset $A \subset X$ and $i : A \hookrightarrow X$ denotes the inclusion. A **deformation retraction** (*Deformationsretraktion*) of X to A is a homotopy $H : I \times X \to X$ such that $H(s, \cdot)|_A = \mathrm{Id}_A$ for every $s \in I$, $H(1, \cdot) = \mathrm{Id}_X$ and $H(0, \cdot) = i \circ f$ for some retraction $f : X \to A$. If a deformation retraction exists, we say that A is a a **deformation retract** (*Deformationsretrakt*) of X.

You should imagine a deformation retraction as a gradual “pulling” of all points in X toward the subset A until eventually all of them end up in A.

Example 10.12. We call $X \subset \mathbb{R}^n$ a **star-shaped domain** (*sternförmige Menge*) if for every $x \in X$, the rescaled vector tx is also in X for every $t \in [0, 1]$. In this case $H(t, x) := tx$ defines a deformation retraction of X to the one-point subset $\{0\}$.

Example 10.13. This is actually a non-example: while the maps $f : S^1 \vee S^1 \to S^1$ in Example 10.8 are retractions, $i \circ f$ in this case is not homotopic to the identity on $S^1 \vee S^1$, so S^1 is not a deformation retract of $S^1 \vee S^1$. We are not yet in a position to prove this, as it will require more knowledge of $\pi_1(S^1 \vee S^1)$ than we presently have, but the necessary results will be proved within the next four lectures. For now, feel free to try to imagine how you might define a homotopy of maps $S^1 \vee S^1 \to S^1 \vee S^1$ that starts with the identity and ends with a retraction collapsing one of the circles. (Keep in mind however that it is not possible, so don't try too hard.)

Example 10.14. The sphere $S^{n-1} \subset \mathbb{R}^n\backslash\{0\}$ is a deformation retract of the punctured Euclidean space. A suitable homotopy $H : I \times (\mathbb{R}^n\backslash\{0\}) \to \mathbb{R}^n\backslash\{0\}$ can be defined by

$$H(t, x) = \frac{x}{t + (1 - t)|x|},$$

which makes $H(1, \cdot)$ the identity map, while $H(0, x) := x/|x|$ retracts $\mathbb{R}^n\backslash\{0\}$ to S^{n-1} and $H(t, x) = x$ for $x \in S^{n-1}$. It is important to observe that no continuous map can be defined in this way with all of $\mathbb{R}^n$ as its domain: the removal of one point changes the topology of $\mathbb{R}^n$ in an essential way that makes the deformation retraction to S^{n-1} possible. (We will later be able to prove that $\mathbb{R}^n$ does not admit any retraction to S^{n-1}. When $n = 2$, this already follows from Theorem 10.10 since $\pi_1(S^1) \cong \mathbb{Z}$ and $\pi_1(\mathbb{R}^2) = 0$.)

Example 10.15. Writing $S^n = \big\{(\mathbf{x}, z) \in \mathbb{R}^n \times \mathbb{R} \;\big|\; |\mathbf{x}|^2 + z^2 = 1\big\}$, define the two "poles" $p_\pm = (0, \pm 1)$. Removing these poles produces a space that can be decomposed into a 1-parameter family of $(n-1)$-spheres, i.e. there is a homeomorphism

$$S^n \backslash \{p_+, p_-\} \xrightarrow{\cong} S^{n-1} \times (-1, 1) : (\mathbf{x}, z) \mapsto \left(\frac{\mathbf{x}}{|\mathbf{x}|}, z\right).$$

If we identify $S^n \backslash \{p_+, p_-\}$ with $S^{n-1} \times (-1, 1)$ in this way, then we see that the "equator" $S^{n-1} \times \{0\} \subset S^n$ is a deformation retract of $S^n \backslash \{p_+, p_-\}$. This follows from the fact that $\{0\}$ is a deformation retract of $(-1, 1)$.

Definition 10.16. A map $f : X \to Y$ is a **homotopy equivalence** (*Homotopieäquivalenz*) if there exists a map $g : Y \to X$ such that $g \circ f$ and $f \circ g$ are each homotopic to the identity map on X and Y respectively. When this exists, we say that g is a **homotopy inverse** (*Homotopieinverse*) of f, and that the spaces X and Y are **homotopy equivalent** (*homotopieäquivalent*). This defines an equivalence relation on topological spaces which we shall denote in these notes by

$$X \underset{h.e.}{\simeq} Y.$$

Exercise 10.17. Verify that homotopy equivalence defines an equivalence relation.

Remark 10.18. The notation "$\underset{h.e.}{\simeq}$" for homotopy equivalence is not universal, and there are several similar but slightly different standards that frequently appear in the literature. This one happens to be my current favorite, but I may change to something else next year.

Example 10.19. A homeomorphism $f : X \to Y$ is obviously also a homotopy equivalence, with homotopy inverse f^{-1}.

Example 10.20. If $H : I \times X \to X$ is a deformation retraction with $H(0, \cdot) = f \circ i$ for a retraction $f : X \to A$, then the inclusion $i : A \hookrightarrow X$ is a homotopy inverse of f, so that both f and i are homotopy equivalences and thus $X \underset{h.e.}{\simeq} A$. Indeed, the retraction condition implies that $f \circ i$ is not just homotopic but also equal to Id_A, and adding the word "deformation" provides the condition $i \circ f \underset{h}{\sim} \mathrm{Id}_X$.

Definition 10.21. We say that a space X is **contractible** (*zusammenziehbar* or *kontrahierbar*) if it is homotopy equivalent to a one-point space.

Remark 10.22. The above definitions imply immediately that any space admitting a deformation retraction to a one-point subset (as in Example 10.12) is contractible. The converse is not quite true. Indeed, suppose $\{x\}$ is a one-point space and $f : X \to \{x\}$ is a homotopy equivalence with homotopy inverse $g : \{x\} \to X$ and a homotopy $H : I \times X \to X$ from Id_X to $g \circ f$. (We do not need to discuss any homotopy of $f \circ g$ since there is only one map $\{x\} \to \{x\}$.) Then if $p := g(x) \in X$, $F : X \to \{p\}$ denotes the constant map at p and $i : \{p\} \hookrightarrow X$ is the inclusion, we have $F \circ i = \mathrm{Id}_{\{p\}}$, and H is a homotopy from Id_X to $i \circ F$. Unfortunately, the definition of homotopy equivalence does not guarantee that this homotopy will satisfy $H(t, p) = p$ for all $t \in I$, so H might not be a deformation retraction in the strict sense of Definition 10.11. It turns out that this distinction matters, but only for fairly strange spaces: see [**Hat02**, p. 18, Exercise 6] for an example of a space that is contractible but does not admit a deformation retraction to any point.

We can now state the main theorem of this lecture.

Theorem 10.23. *If $f : X \to Y$ is a homotopy equivalence with $f(p) = q$, then the induced homomorphism $f_* : \pi_1(X, p) \to \pi_1(Y, q)$ is an isomorphism.*

Since a one-point space contains only one path and therefore has trivial fundamental group, this implies:

Corollary 10.24. *For every contractible space X, $\pi_1(X) = 0$.* □

Proof of Theorem 10.23. Here is a preliminary remark: if you're only half paying attention, then you might reasonably think this theorem follows immediately from Theorem 8.11. Indeed, we stated in that theorem that the homomorphism $f_* : \pi_1(X,p) \to \pi_1(Y,q)$ depends only on the pointed homotopy class of f, and the same is of course true of the compositions $g \circ f$ and $f \circ g$, which ought to make $g_* \circ f_*$ and $f_* \circ g_*$ both the identity if $g \circ f$ and $f \circ g$ are homotopic to the identity. The problem however is that we are not paying attention to the base point: the definition of homotopy equivalence never mentions any base point and says "homotopy" rather than "pointed homotopy," while in Theorem 8.11, maps and homotopies are always required to preserve base points. In particular, if $f(p) = q$ and $g : Y \to X$ is a homotopy inverse of f, then there is no reason to expect $g(q) = p$, in which case $g_* : \pi_1(Y,q) \to \pi_1(X,g(q))$ cannot be an inverse of $f_* : \pi_1(X,p) \to \pi_1(Y,q)$, as its target is not even the same group as the domain of f_*. The main content of the following proof is an argument to cope with this annoying detail.

With that out of the way, assume $f : X \to Y$ is a map with homotopy inverse $g : Y \to X$, satisfying $f(p) = q$ and $g(q) = r$, so we have a sequence of pointed maps

$$(X,p) \xrightarrow{f} (Y,q) \xrightarrow{g} (X,r)$$

and induced homomorphisms

$$\pi_1(X,p) \xrightarrow{f_*} \pi_1(Y,q) \xrightarrow{g_*} \pi_1(X,r). \tag{10.2}$$

By assumption there exists a homotopy $H : I \times X \to X$, which we shall write as a 1-parameter family of maps

$$h_s := H(s,\cdot) : X \to X \quad \text{for} \quad s \in I,$$

satisfying $h_0 = \mathrm{Id}_X$ and $h_1 = g \circ f$. We can therefore define a path $p \overset{\gamma}{\rightsquigarrow} r$ by

$$\gamma(t) := h_t(p),$$

and by Theorem 9.1, this gives rise to an isomorphism

$$\Phi_\gamma : \pi_1(X,r) \to \pi_1(X,p) : [\alpha] \mapsto [\gamma \cdot \alpha \cdot \gamma^{-1}].$$

We claim that the diagram

$$\begin{tikzcd}
\pi_1(X,p) \arrow[r, "f_*"] \arrow[dr, "\Phi_\gamma^{-1}"'] & \pi_1(Y,q) \arrow[d, "g_*"] \\
 & \pi_1(X,r)
\end{tikzcd}$$

commutes, or equivalently, $\Phi_\gamma \circ g_* \circ f_*$ is the identity map on $\pi_1(X,p)$. Given a loop $p \overset{\alpha}{\rightsquigarrow} p$, the element $\Phi_\gamma \circ g_* \circ f_*[\alpha] = \Phi_\gamma \circ (g \circ f)_*[\alpha]$ is represented by $\gamma \cdot (g \circ f \circ \alpha) \cdot \gamma^{-1}$, so we need to show that the latter is homotopic with fixed end points to α. A precise formula for such a homotopy is provided by the following 1-parameter family of loops: for $s \in I$, let

$$\alpha_s := \gamma_s \cdot (h_s \circ \alpha) \cdot \gamma_s^{-1},$$

where $p \overset{\gamma_s}{\rightsquigarrow} \gamma(s)$ denotes the path $\gamma_s(t) := \gamma(st)$. (For a visualization of what this homotopy is actually doing, I recommend the picture on page 37 of [**Hat02**].) This proves the claim, and since Φ_γ is an isomorphism, it implies that $g_* \circ f_* = \Phi_\gamma^{-1}$ is also an isomorphism, from which we deduce that f_* is injective and g_* is surjective.

The preceding argument was based on the assumption that $g \circ f : X \to X$ is homotopic to the identity. We have not yet used the assumption that $f \circ g : Y \to Y$ is also homotopic to the identity, but we can use it now to carry out the same argument again with the roles of f and g reversed. The conclusion is that $f_* \circ g_*$ is also an isomorphism, implying g_* is injective and f_* is surjective. We conclude that f_* and g_* are in fact both isomorphisms. □

EXAMPLE 10.25. Here are some examples of contractible spaces, which therefore have isomorphic (trivial) fundamental groups even though they are not all homeomorphic: $\mathbb{R}^n$, $\mathbb{D}^n$ (not homeomorphic to $\mathbb{R}^n$ since it is compact), any convex subset or star-shaped domain in $\mathbb{R}^n$ as in Example 10.12. A quite different type of example comes from *graph theory*: a **graph** is a combinatorial object consisting of a set V (called the **vertices**) and a set E whose elements (the **edges**) are unordered pairs of vertices. A graph is typically represented by depicting the vertices as points and the edges $\{x, y\} \in E$ as curves connecting the corresponding vertices x and y to each other. One can thus naturally view a graph as a topological space in which each vertex is a point and each edge is a subset homeomorphic to $[0, 1]$ (possibly with its end points identified if its two vertices are the same one). A graph is called a **tree** if there is exactly one path (up to parametrization) connecting any two of its vertices. It is not hard to show that any finite graph with this property is a contractible space: pick your favorite vertex $v \in V$, draw the unique path from v to every other vertex, then define a deformation retraction to v by pulling everything back along these paths.

EXAMPLE 10.26. Viewing S^1 as the unit circle in $\mathbb{C}$, associate to each $z \in \mathbb{C}$ the loop $\gamma_z : S^1 \hookrightarrow \mathbb{C}\backslash\{z\} : e^{i\theta} \mapsto z + e^{i\theta}$. Since these are pointed maps $(S^1, 1) \to (\mathbb{C}\backslash\{z\}, z + 1)$, they represent elements $[\gamma_z] \in \pi_1(\mathbb{C}\backslash\{z\}, z + 1)$. We claim in fact that this group is isomorphic to $\mathbb{Z}$, and that $[\gamma_z]$ generates it. The proof is mainly the observation that $\gamma_z(S^1)$ is a deformation retract of $\mathbb{C}\backslash\{z\}$, by a construction analogous to Example 10.14, hence γ_z is a homotopy equivalence and therefore induces an isomorphism $\pi_1(S^1, 1) \to \pi_1(\mathbb{C}\backslash\{z\}, z + 1)$. Since the identity map $(S^1, 1) \to (S^1, 1)$ represents a generator of $\pi_1(S^1, 1)$, composing this with γ_z now represents a generator of $\pi_1(\mathbb{C}\backslash\{z\}, z + 1)$ as claimed.

EXERCISE 10.27. For a point $z \in \mathbb{C}$ and a continuous map $\gamma : [0, 1] \to \mathbb{C}\backslash\{z\}$ with $\gamma(0) = \gamma(1)$, one defines the **winding number** of γ about z as

$$\mathrm{wind}(\gamma; z) = \theta(1) - \theta(0) \in \mathbb{Z}$$

where $\theta : [0, 1] \to \mathbb{R}$ is any choice of continuous function such that

$$\gamma(t) = z + r(t)e^{2\pi i\theta(t)}$$

for some function $r : [0, 1] \to (0, \infty)$. Notice that since $\gamma(t) \neq z$ for all t, the function $r(t)$ is uniquely determined, and requiring $\theta(t)$ to be continuous makes it unique up to the addition of a constant integer, hence $\theta(1) - \theta(0)$ depends only on the path γ and not on any additional choices. One of the fundamental facts about winding numbers is their important role in the computation of $\pi_1(S^1)$: as we saw in Example 9.12, viewing S^1 as $\{z \in \mathbb{C} \mid |z| = 1\}$, the map

$$\pi_1(S^1, 1) \to \mathbb{Z} : [\gamma] \mapsto \mathrm{wind}(\gamma; 0)$$

is an isomorphism to the abelian group $(\mathbb{Z}, +)$. Assume in the following that $\Omega \subset \mathbb{C}$ is an open set and $f : \Omega \to \mathbb{C}$ is a continuous function.

(a) Suppose $f(z) = w$ and $w \notin f(\mathcal{U}\backslash\{z\})$ for some neighborhood $\mathcal{U} \subset \Omega$ of z. This implies that the loop $f \circ \gamma_\epsilon$ for $\gamma_\epsilon : [0, 1] \to \Omega : t \mapsto z + \epsilon e^{2\pi i t}$ has image in $\mathbb{C}\backslash\{w\}$ for all $\epsilon > 0$ sufficiently small, hence $\mathrm{wind}(f \circ \gamma_\epsilon; w)$ is well defined. Show that for some $\epsilon_0 > 0$, $\mathrm{wind}(f \circ \gamma_\epsilon; w)$ does not depend on ϵ as long as $0 < \epsilon \leqslant \epsilon_0$.

(b) Show that if the ball $B_r(z_0)$ of radius $r > 0$ about $z_0 \in \Omega$ has its closure contained in Ω, and the loop $\gamma(t) = z_0 + re^{2\pi it}$ satisfies $\mathrm{wind}(f \circ \gamma; w) \neq 0$ for some $w \in \mathbb{C}$, then there exists $z \in B_r(z_0)$ with $f(z) = w$.
Hint: Recall that if we regard elements of $\pi_1(X, p)$ as pointed homotopy classes of maps $S^1 \to X$, then such a map represents the identity in $\pi_1(X, p)$ if and only if it admits a continuous extension to a map $\mathbb{D}^2 \to X$. Define X in the present case to be $\mathbb{C}\backslash\{w\}$.

(c) Prove the Fundamental Theorem of Algebra: every nonconstant complex polynomial has a root.
Hint: Consider loops $\gamma(t) = Re^{2\pi it}$ with $R > 0$ large.

(d) We call $z_0 \in \Omega$ an **isolated zero** of $f : \Omega \to \mathbb{C}$ if $f(z_0) = 0$ but $0 \notin f(\mathcal{U}\backslash\{z_0\})$ for some neighborhood $\mathcal{U} \subset \Omega$ of z_0. Let us say that such a zero has **order** $k \in \mathbb{Z}$ if $\mathrm{wind}(f \circ \gamma_\epsilon; 0) = k$ for $\gamma_\epsilon(t) = z_0 + \epsilon e^{2\pi it}$ and $\epsilon > 0$ small (recall from part (a) that this does not depend on the choice of ϵ if it is small enough). Show that if $k \neq 0$, then for any neighborhood $\mathcal{U} \subset \Omega$ of z_0, there exists $\delta > 0$ such that every continuous function $g : \Omega \to \mathbb{C}$ satisfying $|f - g| < \delta$ everywhere has a zero somewhere in $\mathcal{U}$.

(e) Find an example of the situation in part (d) with $k = 0$ such that f admits arbitrarily close perturbations g that have no zeroes in some fixed neighborhood of $\mathcal{U}$.
Hint: Write f as a continuous function of x and y where $x + iy \in \Omega$. You will not be able to find an example for which f is holomorphic—they do not exist!

General advice: Throughout this problem, it is important to remember that $\mathbb{C}\backslash\{w\}$ is homotopy equivalent to S^1 for every $w \in \mathbb{C}$. Thus all questions about $\pi_1(\mathbb{C}\backslash\{w\})$ can be reduced to questions about $\pi_1(S^1)$.

11. The easy part of van Kampen's theorem

The main question of this lecture is the following: If X is the union of two subsets $A \cup B$ and we know both $\pi_1(A)$ and $\pi_1(B)$, what can we say about $\pi_1(X)$?

EXAMPLE 11.1. The sphere S^n can be viewed as the union of two subsets A and B that are both homeomorphic to $\mathbb{D}^n$, e.g. when $n = 2$, we would take the northern and southern "hemispheres" of the globe. Since $\mathbb{D}^n$ is contractible, $\pi_1(A) = \pi_1(B) = 0$. We will see below that this is almost enough information to compute $\pi_1(S^n)$.

The next lemma is the "easy" first half of an important result about fundamental groups known as the *Seifert-van Kampen theorem*, or often simply *van Kampen's theorem*. The much more powerful "hard" part of the theorem will be dealt with in the two subsequent lectures, though the easy part already has several impressive applications. We will state it here in somewhat greater generality than is needed for most applications: on first reading, you are free to replace the arbitrary open covering $X = \bigcup_{\alpha \in J} A_\alpha$ with a covering by *two* open subsets $X = A \cup B$, which will be the situation in all of the examples below.

LEMMA 11.2. *Suppose $X = \bigcup_{\alpha \in J} A_\alpha$ for a collection of open subsets $\{A_\alpha \subset X\}_{\alpha \in J}$ satisfying the following conditions:*

(1) A_α is path-connected for every $\alpha \in J$;
(2) $A_\alpha \cap A_\beta$ is path-connected for every pair $\alpha, \beta \in J$;
(3) $\bigcap_{\alpha \in J} A_\alpha \neq \varnothing$.

Let $A_\alpha \overset{i_\alpha}{\hookrightarrow} X$ denote the natural inclusion maps. Then for any base point $p \in \bigcap_{\alpha \in J} A_\alpha$, $\pi_1(X, p)$ is generated by the subgroups

$$(i_\alpha)_* (\pi_1(A_\alpha, p)) \subset \pi_1(X, p),$$

i.e. every element of $\pi_1(X,p)$ *is a product of elements of the form* $(i_\alpha)_*[\gamma]$ *for some* $\alpha \in J$ *and* $[\gamma] \in \pi_1(A_\alpha, p)$.

Before proving the lemma, let's look at several more examples, starting with a rehash of Example 11.1 above.

Example 11.3. Denote points in the unit sphere S^n by $(\mathbf{x}, z) \in \mathbb{R}^n \times \mathbb{R}$ such that $|\mathbf{x}|^2 + z^2 = 1$, and define the open subsets

$$A := \{z > -\epsilon\} \subset S^n, \qquad B := \{z < \epsilon\} \subset S^n$$

for some $\epsilon > 0$ small. Then $A \cong B \cong \mathbb{R}^n$, so both have trivial fundamental group. Moreover, $A \cap B \cong S^{n-1} \times (-\epsilon, \epsilon)$ is path-connected if $n \geqslant 2$. (Note that this is not true if $n = 1$: the 0-sphere S^0 is just the set of two points $\{1, -1\} \subset \mathbb{R}$, so it is not path-connected.) The lemma therefore implies that for any $p \in A \cap B$, $\pi_1(S^n, p)$ is generated by images of homomorphisms into $\pi_1(S^n, p)$ from the groups $\pi_1(A, p)$ and $\pi_1(B, p)$, both of which are trivial, therefore $\pi_1(S^n, p)$ is trivial.

We just proved:

Corollary 11.4. *For all* $n \geqslant 2$, S^n *is simply connected.* ☐

Here is an easy application:

Theorem 11.5. *For every* $n \geqslant 3$, $\mathbb{R}^2$ *is not homeomorphic to* $\mathbb{R}^n$.

Proof. The complement of one point in $\mathbb{R}^n$ is homotopy eqivalent to S^{n-1}, thus $\pi_1(\mathbb{R}^n \backslash \{\text{pt}\}) \cong \pi_1(S^{n-1}) = 0$ if $n \geqslant 3$, while $\pi_1(\mathbb{R}^2 \backslash \{\text{pt}\}) \cong \pi_1(S^1) \cong \mathbb{Z}$. It follows that $\mathbb{R}^2 \backslash \{\text{pt}\}$ and $\mathbb{R}^n \backslash \{\text{pt}\}$ for $n \geqslant 3$ are not homeomorphic, hence neither are $\mathbb{R}^2$ and $\mathbb{R}^n$. ☐

A wider class of examples comes from the following general construction known as *gluing* of spaces. Assume X, Y and A are spaces and we have inclusions[9]

$$i_X : A \hookrightarrow X, \qquad i_Y : A \hookrightarrow Y.$$

We then define the space

$$X \cup_A Y := (X \amalg Y)/\sim$$

where the equivalence relation identifies $i_X(a) \in X$ with $i_Y(a) \in Y$ for every $a \in A$. As usual in such constructions, we assign to $X \amalg Y$ the disjoint union topology and then give $X \cup_A Y$ the quotient topology. We say that $X \cup_A Y$ is the space obtained by **gluing X to Y along A.** Note that we can regard X and Y both as subspaces of $X \cup_A Y$, and their intersection is a subspace homeomorphic to A. The wedge sum of two spaces (see Example 10.3) is the special case of this construction where A is a single point. (The notation is slightly non-ideal since $X \cup_A Y$ depends on the inclusions of A into X and Y, not just on the three spaces themselves, but in most interesting examples the inclusions are obvious, so the notation is easy to interpret.)

Example 11.6. If $X = Y = \mathbb{D}^n$ and $A = S^{n-1}$ is included in both as the boundary $\partial\mathbb{D}^n$, then the descriptions of S^n in Examples 11.1 and 11.3 translates into

$$\mathbb{D}^n \cup_{S^{n-1}} \mathbb{D}^n \cong S^n.$$

---

[9] The technical meaning of the word **inclusion** in this context is a map $A \hookrightarrow X$ which is injective and is a homeomorphism onto its image (with the subspace topology). Such a map is also sometimes called a **topological embedding**.

EXAMPLE 11.7. In Example 1.2 we gave a description of $\mathbb{RP}^2$ as the space obtained by gluing a disk $\mathbb{D}^2$ to a Möbius strip

$$\mathbb{M} := \left\{ (e^{i\theta}, t\cos(\theta/2), t\sin(\theta/2)) \in S^1 \times \mathbb{R}^2 \;\middle|\; e^{i\theta} \in S^1,\ t \in [-1,1] \right\}$$

along their boundaries, which are both homeomorphic to S^1. Choose a particular inclusion of S^1 as the boundary of $\mathbb{M}$, e.g.

$$S^1 \hookrightarrow \mathbb{M} : e^{i\theta} \mapsto (e^{2i\theta}, \cos(\theta), \sin(\theta)).$$

Then our picture of $\mathbb{RP}^2$ can be expressed succinctly as

$$\mathbb{RP}^2 \cong \mathbb{D}^2 \cup_{S^1} \mathbb{M}.$$

Lemma 11.2 can now be applied to this as follows. There is an obvious deformation retraction of $\mathbb{M}$ to the "central" circle $S^1 \times \{0\} \subset \mathbb{M}$, defined via the homotopy

$$H : I \times \mathbb{M} \to \mathbb{M} : (s, (e^{i\theta}, t\cos(\theta/2), t\sin(\theta/2))) \mapsto (e^{i\theta}, st\cos(\theta/2), st\sin(\theta/2)),$$

thus $\mathbb{M} \underset{h.e.}{\simeq} S^1$. The gluing construction allows us to view both $\mathbb{D}^2$ and $\mathbb{M}$ as subsets of $\mathbb{RP}^2$, but they are not *open* subsets as required by the lemma. This can easily be fixed by slightly expanding both of them. Concretely, by adding a neighborhood of $\partial\mathbb{M}$ in $\mathbb{M}$ to $\mathbb{D}^2$, we obtain an open neighborhood $A \subset \mathbb{RP}^2$ of $\mathbb{D}^2$ that is homeomorphic to an open disk, and similarly, adding a neighborhood of $\partial\mathbb{D}^2$ in $\mathbb{D}^2$ to $\mathbb{M}$ gives an open neighborhood $B \subset \mathbb{RP}^2$ of $\mathbb{M}$ that admits a deformation retraction to $\mathbb{M}$ and thus also to the central circle $S^1 \times \{0\} \subset \mathbb{M}$. We now have

$$\pi_1(A) \cong \pi_1(\mathring{\mathbb{D}}^2) = 0 \quad \text{and} \quad \pi_1(B) \cong \pi_1(\mathbb{M}) \cong \pi_1(S^1) \cong \mathbb{Z},$$

and notice also that A and B are both path connected, and so is $A \cap B$ since we can arrange for the latter to be homeomorphic to $S^1 \times (-1,1)$, i.e. it is the union of an annular neighborhood of $\partial\mathbb{D}^2$ in $\mathbb{D}^2$ with another annular neighborhood of $\partial\mathbb{M}$ in $\mathbb{M}$. The lemma thus implies that for any $p \in A \cap B$, $\pi_1(\mathbb{RP}^2, p)$ is generated by the element $i^B_*[\gamma] \in \pi_1(\mathbb{RP}^2, p)$, where $i^B : B \hookrightarrow \mathbb{RP}^2$ is the inclusion and $\gamma : (S^1, 1) \to (B, p)$ is any loop such that $[\gamma]$ generates $\pi_1(B, p) \cong \mathbb{Z}$. In light of the deformation retraction to the central circle, the inclusion of that circle into B induces an isomorphism of fundamental groups, thus we can take γ to be the obvious inclusion of S^1 into B as the central circle:

$$\begin{aligned} \gamma : S^1 &\xrightarrow{\cong} S^1 \times \{0\} \subset \mathbb{M} \subset \mathbb{RP}^2, \\ e^{i\theta} &\mapsto (e^{i\theta}, 0). \end{aligned} \tag{11.1}$$

The conclusion is that if we regard γ in this way as a loop in $\mathbb{RP}^2$, then $[\gamma]$ generates $\pi_1(\mathbb{RP}^2, p)$. The loop γ is not hard to visualize if you translate from our picture of $\mathbb{RP}^2$ as $\mathbb{D}^2 \cup_{S^1} \mathbb{M}$ back to the usual definition of $\mathbb{RP}^2$ as a quotient of S^2 (see Example 1.2): in the latter picture you can realize γ as a path along the equator of S^2 that goes exactly halfway around. Note that this is not a loop in S^2, but it becomes a loop when you project it to $\mathbb{RP}^2$ since its starting and end point are antipodal.

A word of caution is in order: we have not yet actually computed $\pi_1(\mathbb{RP}^2)$, we have only shown that every element in $\pi_1(\mathbb{RP}^2)$ is a power of a single element $[\gamma]$. It is still possible that $\pi_1(\mathbb{RP}^2)$ is trivial because γ is contractible—this will turn out not to be the case, but we are not in a position to prove it just yet. We can say one more thing, however: $[\gamma]^2$ is the identity element in $\pi_1(\mathbb{RP}^2, p)$. Indeed, $[\gamma]^2$ is represented by the concatenation of γ with itself, which can also be realized as the projection through $S^2 \xrightarrow{\pi} \mathbb{RP}^2$ of a path that goes *all the way* around the equator in S^2, i.e. it is the concatenation of two paths that go halfway around. But if $\alpha : S^1 \to S^2$ parametrizes this loop around the equator, then there is obviously an extension of α to a map $u : \mathbb{D}^2 \to S^2$ satisfying $u|_{\partial\mathbb{D}^2} = \alpha$, namely the inclusion of either the northern or southern hemisphere of S^2.

The map $\pi \circ u : \mathbb{D}^2 \to \mathbb{RP}^2$ is then an extension over the disk of our loop representing $[\gamma]^2$, which proves via Theorem 9.4 that $[\gamma]^2$ is trivial. This proves that $\pi_1(\mathbb{RP}^2)$ is either the trivial group or is isomorphic to $\mathbb{Z}_2$; we will see that it is the latter when we prove that the generator $[\gamma]$ is nontrivial.

Here is another pair of general constructions that produce many more examples.

Definition 11.8. Given a space X, the **cone** (*Kegel*) of X is the space

$$CX := (X \times I)/(X \times \{1\}).$$

The single point in CX represented by $(x, 1)$ for every $x \in X$ is sometimes called the "summit" or "node" of the cone.

Exercise 11.9. Show that CS^{n-1} is homeomorphic to $\mathbb{D}^n$.

Lemma 11.10. *For every space X, the cone CX is contractible.*

Proof. There is an obvious deformation retraction of $X \times I$ to $X \times \{1\}$ defined by pushing every $(x, t) \in X \times I$ upward in the t-coordinate. Writing down this same deformation retraction on the quotient $(X \times I)/(X \times \{1\})$, the result is that everything gets pushed to a single point, the summit of the cone. □

Definition 11.11. Given a space X, the **suspension** (*Einhängung*) of X is the space

$$SX := C_+X \cup_{X\times\{0\}} C_-X,$$

where $C_+X := CX$ as above, and C_-X is the "reversed" cone $(X \times [-1, 0])/(X \times \{-1\})$. Equivalently, the suspension can be written as

$$SX = (X \times [-1, 1])/\sim$$

where $(x, 1) \sim (y, 1)$ and $(x, -1) \sim (y, -1)$ for every $x, y \in X$.

Exercise 11.12. Show that $SS^{n-1} \cong S^n$.

We can now generalize the result that $\pi_1(S^n) = 0$ for $n \geqslant 2$ as follows.

Theorem 11.13. *If X is path-connected, then its suspension SX is simply connected.*

Proof. We define $A, B \subset SX$ to be open neighborhoods of C_+X and C_-X respectively, e.g.

$$A := (X \times (-\epsilon, 1])/(X \times \{1\}), \qquad B := (X \times [-1, \epsilon))/(X \times \{-1\})$$

for any $\epsilon \in (0, 1)$. The subspaces are both contractible for the same reason that C_+X and C_-X are: one can define deformation retractions to a point by pushing upward in A and downward in B. Moreover, $A \cap B = X \times (-\epsilon, \epsilon)$ is path-connected if and only if X is path-connected, and in that case, Lemma 11.2 implies that $\pi_1(SX)$ is generated by the images of homomorphisms from $\pi_1(A)$ and $\pi_1(B)$, both of which are trivial, therefore $\pi_1(SX)$ is trivial. □

Let us finally prove the lemma.

Proof of Lemma 11.2. We assume $X = \bigcup_{\alpha\in J} A_\alpha$ and $p \in \bigcap_{\alpha\in J} A_\alpha$, where the sets $A_\alpha \subset X$ are open and path-connected, and $A_\alpha \cap A_\beta$ is also path-connected for every pair $\alpha, \beta \in J$. What we need to show is that every loop $p \overset{\gamma}{\rightsquigarrow} p$ in X is homotopic with fixed end points to a concatenation of finitely many loops based at p that are each contained in one of the subsets A_α. To start with, observe that since $\gamma : I \to X$ is continuous, $I_\alpha := \gamma^{-1}(A_\alpha)$ is an open subset of I for every α, and is therefore a union of open subintervals of I.[10] The union of all these open subintervals for all

---

[10] Remember that since sets like $[0, \epsilon) \subset I$ that include an end point are open subsets of I, they are included in the term "open subinterval of I".

$\alpha \in J$ thus forms an open covering of I, which has a finite subcovering since I is compact, giving rise to a finite collection of open subintervals

$$I = I_1 \cup \ldots \cup I_N$$

such that for each $j = 1, \ldots, N$, $\gamma(I_j) \subset A_{\alpha_j}$ for some $\alpha_j \in J$. After relabeling the α_j's if necessary, we can then find a finite increasing sequence

$$0 =: t_0 < t_1 < \ldots < t_{N-1} < t_N := 1$$

such that $\gamma([t_{j-1}, t_j]) \subset A_{\alpha_j}$ for each $j = 1, \ldots, N$. In particular, for $j = 1, \ldots, N-1$, each $\gamma(t_j)$ lies in both A_{α_j} and $A_{\alpha_{j+1}}$. The intersection of these two sets is path-connected by assumption, so choose a path β_j in $A_{\alpha_j} \cap A_{\alpha_{j+1}}$ from $\gamma(t_j)$ to the base point p. Then if we write $\gamma_j := \gamma|_{[t_{j-1}, t_j]}$ and reparametrize each of these paths to define them on the usual interval I, we have

$$\gamma = \gamma_1 \cdot \ldots \cdot \gamma_N \underset{h+}{\sim} \gamma_1 \cdot \beta_1 \cdot \beta_1^{-1} \cdot \gamma_2 \cdot \beta_2 \cdot \beta_2^{-1} \cdot \ldots \cdot \beta_{N-2} \cdot \beta_{N-2}^{-1} \cdot \gamma_{N-1} \cdot \beta_{N-1} \cdot \beta_{N-1}^{-1} \cdot \gamma_N.$$

The latter is the concatenation we were looking for since $\gamma_1 \cdot \beta_1$ is a loop from p to itself in A_{α_1}, $\beta_1^{-1} \cdot \gamma_2 \cdot \beta_2$ is a loop from p to itself in A_{α_2}, and so forth up to $\beta_{N-2}^{-1} \cdot \gamma_{N-1} \cdot \beta_{N-1}$ in $A_{\alpha_{N-1}}$ and $\beta_{N-1}^{-1} \cdot \gamma_N$ in A_{α_N}. ☐

To conclude this lecture, we would like to restate Lemma 11.2 in more precise terms. This requires a few notions from combinatorial group theory.

Definition 11.14. Suppose $\{G_\alpha\}_{\alpha \in J}$ is a collection of groups, with the identity element in each denoted by $e_\alpha \in G_\alpha$. For any integer $N \geqslant 0$, an ordered set $b_1 b_2 \ldots b_N$ together with a corresponding ordered set $\alpha_1, \alpha_2, \ldots, \alpha_N \in J$ is called a **word in** $\{G_\alpha\}_{\alpha \in J}$ if $b_i \in G_{\alpha_i}$ for each $i = 1, \ldots, N$. Informally, we call the elements of the sequence *letters*, and denote the word by $b_1 \ldots b_N$ even though, strictly speaking, the set of indices $\alpha_1, \ldots, \alpha_N \in J$ is also part of the data defining the word.[11] Note that this definition includes the so-called *empty word*, with $N = 0$, i.e. the word with no letters. A word $a_1 \ldots a_N$ is called a **reduced word** if:

- none of the letters b_i are the identity element $e_{\alpha_i} \in G_{\alpha_i}$ in the corresponding group, and
- no two adjacent letters b_i and b_{i+1} satisfy $\alpha_i = \alpha_{i+1}$, i.e. the groups that appear in adjacent positions are distinct.

Note that the empty word trivially satisfies both conditions, thus it is a reduced word.

There is an obvious map called **reduction** from the set of all words to the set of all reduced words: it acts on a given word $b_1 \ldots b_N$ by replacing all adjacent pairs $b_i b_{i+1}$ with their product in G_α whenever $\alpha_i = \alpha_{i+1} = \alpha$, and removing all e_α's.

Definition 11.15. The **free product** (*freies Produkt*) $\circledast_{\alpha \in J} G_\alpha$ of a collection of groups $\{G_\alpha\}_{\alpha \in J}$ is defined as the set of all reduced words in $\{G_\alpha\}_{\alpha \in J}$. The product of two reduced words $w = b_1 \ldots b_N$ and $w' = b'_1 \ldots b'_{N'}$ in this group is defined to be the reduction of the concatenated word $ww' = b_1 \ldots b_N b'_1 \ldots b'_{N'}$. The identity element is the empty word, and will be denoted by

$$e \in \underset{\alpha \in J}{\circledast} G_\alpha.$$

We will typically deal with collections of only finitely many groups $G_1, \ldots, G_N$, in which case the free product is usually denoted by

$$G_1 * \ldots * G_N.$$

---

[11] This is important to remember in case some G_α and G_β contain common elements for $\alpha \neq \beta$, e.g. if they are both subgroups of a single larger group. If not, then this detail is safe to ignore and the notation $b_1 \ldots b_N$ for a word is completely unambiguous.

In general, this is an enormous group, e.g. it is always infinite if there are at least two nontrivial groups in the collection, no matter how small those groups are. It is also always nonabelian in those cases. Let us see some examples.

Example 11.16. Consider two copies of the same group $G = H = \mathbb{Z}_2$, with the unique nontrivial elements of G and H denoted by $a \in G$ and $b \in H$. Then $G * H$ consists of all possible reduced words built out of these two letters, plus the empty word e, so

$$\mathbb{Z}_2 * \mathbb{Z}_2 \cong G * H = \{e, a, b, ab, ba, aba, bab, abab, baba, \dots\} .$$

For an example of how multiplication in $\mathbb{Z}_2 * \mathbb{Z}_2$ works, the product of aba and ab is a, i.e. this is the result of reducing the unreduced word $abaab$ since aa and bb are both identity elements.

Example 11.17. Let $G = \mathbb{Z}$ with a generator denoted by $a \in G$, and $H = \mathbb{Z}_2$ with nontrivial element b. If we write G as a multiplicative group so that its elements are all of the form a^p for $p \in \mathbb{Z}$, then

$$\mathbb{Z} * \mathbb{Z}_2 \cong G * H = \left\{e, a^p, b, a^p b, ba^p, a^p ba^q, ba^p ba^q, a^p ba^q ba^r, \dots \;\middle|\; p, q, r, \dots \in \mathbb{Z}\right\} .$$

For an example of a product, $a^p ba^r$ times $a^{-1}b$ gives $a^p ba^{r-1}b$.

With this terminology understood, here is what we actually proved when we proved Lemma 11.2.

Lemma 11.18. *Given* $X = \bigcup_{\alpha \in J} A_\alpha$ *and* $p \in \bigcap_{\alpha \in J} A_\alpha$ *as in Lemma 11.2, there exists a natural group homomorphism*

$$\mathop{\Large *}_{\alpha \in J} \pi_1(A_\alpha, p) \xrightarrow{\Phi} \pi_1(X, p)$$

sending each reduced word $[\gamma_1] \dots [\gamma_N] \in *_{\alpha \in J} \pi_1(A_\alpha, p)$ *with* $[\gamma_i] \in \pi_1(A_{\alpha_i}, p)$ *to the concatenation* $[\gamma_1 \cdot \ldots \cdot \gamma_N] \in \pi_1(X, p)$, *and* Φ *is surjective.* □

The existence of the homomorphism Φ is an easy and purely algebraic fact, which we'll expand on a bit in the next lecture. The truly nontrivial statement here is that Φ is surjective. If we can now identify the kernel of Φ, then Φ descends to an isomorphism from the quotient of the free product by $\ker \Phi$ to $\pi_1(X, p)$, and we will thus have a formula for $\pi_1(X, p)$. Identifying the kernel and then using the resulting formula in applications will be our main topic for the next two lectures.

12. Normal subgroups, generators and relations

Before stating the general version of the Seifert-van Kampen theorem, we need to collect a few more useful algebraic facts about groups and the free product. Recall from the previous lecture that the free product $*_{\alpha \in J} G_\alpha$ of an arbitrary collection of groups $\{G_\alpha\}_{\alpha \in J}$ is defined to consist of all so-called *reduced words* $g_1 \dots g_N$ in which each "letter" g_i is an element of one of the groups G_{α_i}, and the choice of $\alpha_i \in J$ such that $g_i \in G_{\alpha_i}$ for each $i = 1, \dots, N$ is considered part of the data defining the word.[12] The word "reduced" means that the sequence of letters in the word cannot be simplified by computing products in any of the individual groups, hence no consecutive letters $g_i g_{i+1}$ with $\alpha_i = \alpha_{i+1} =: \alpha$ appear—if such a pair appeared then it could be replaced by a single letter formed from the product $g_i g_{i+1} \in G_\alpha$—and similarly, none of the letters is the identity element in any of the groups. Products in $*_{\alpha \in J} G_\alpha$ are formed by concatenating words and then

---

[12] This latter detail is unimportant if the groups G_α are all disjoint sets in the first place, but if any of them have elements in common, e.g. if some G_α and G_β for $\alpha \neq \beta$ are copies of the same group, then we regard them as *separate* copies and always keep track of which letter belongs to which copy. The idea is somewhat analogous to constructing the disjoint union $\coprod_{\alpha \in J} X_\alpha$ of sets, in which X_β and X_γ for $\beta \neq \gamma$ always become disjoint subsets of $\coprod_{\alpha \in J} X_\alpha$, even if they are originally defined as the same set, e.g. $\mathbb{R} \amalg \mathbb{R}$ is by definition two disjoint copies of $\mathbb{R}$, which is different from the ordinary union $\mathbb{R} \cup \mathbb{R} = \mathbb{R}$.

reducing them if necessary, so for example, if G and H are two groups containing elements $g \in G$ and $h, k \in H$, then the product of the reduced words $gh \in G * H$ and $h^{-1}k \in G * H$ is

$$(gh)(h^{-1}k) = gk \in G * H,$$

since the concatenated word $ghh^{-1}k$ can be reduced by replacing hh^{-1} with the identity element $e \in H$ and then removing e from the word. The identity element in $\ast_{\alpha \in J} G_\alpha$ itself is the so-called "empty" word, with zero letters, which we will usually denote by e; there should be no danger of confusing this with the identity elements of the individual groups G_α, since they never appear in reduced words.

The following result is easy to prove directly from the definitions.

Proposition 12.1. *Assume $\{G_\alpha\}_{\alpha \in J}$ is a collection of groups. Then:*

(1) For each $\alpha \in J$, the free product $\ast_{\beta \in J} G_\beta$ contains a distinguished subgroup isomorphic to G_α: it consists of the empty word plus all reduced words of exactly one letter which is in G_α.

(2) If we regard each G_α as a subgroup of $\ast_{\gamma \in J} G_\gamma$ as described above, then for every $\alpha, \beta \in J$ with $\alpha \neq \beta$, the intersection $G_\alpha \cap G_\beta$ in $\ast_{\gamma \in J} G_\gamma$ consists only of the identity element e (i.e. the empty word), and any two nontrivial elements $g \in G_\alpha$ and $h \in G_\beta$ satisfy $gh \neq hg$ in $\ast_{\gamma \in J} G_\gamma$.

(3) For any group H with a collection of homomorphisms $\{\Phi_\alpha : G_\alpha \to H\}_{\alpha \in J}$, there exists a unique homomorphism

$$\Phi : \mathop{\ast}_{\alpha \in J} G_\alpha \to H$$

whose restriction to each of the subgroups $G_\alpha \subset \ast_{\beta \in J} G_\beta$ is Φ_α.

The third item in this list deserves brief comment: the homomorphism $\Phi : \ast_{\alpha \in J} G_\alpha \to H$ exists and is unique because every element of $\ast_{\alpha \in J} G_\alpha$ is uniquely expressible as a reduced word $g_1 \dots g_N$ with $g_i \in G_{\alpha_i}$ for some specified $\alpha_1, \dots, \alpha_N \in J$, hence the definition of Φ can only be

$$\Phi(g_1 \dots g_N) = \Phi_{\alpha_1}(g_1) \dots \Phi_{\alpha_N}(g_N) \in H.$$

It is similarly straightfoward to verify that Φ by this definition is a homomorphism.

Remark 12.2. In Lemma 11.18 at the end of the previous lecture the homomorphism

$$\mathop{\ast}_{\alpha \in J} \pi_1(A_\alpha, p) \xrightarrow{\Phi} \pi_1(X, p) \tag{12.1}$$

is determined as in the proposition above by the homomorphisms $(i_\alpha)_* : \pi_1(A_\alpha, p) \to \pi_1(X, p)$ induced by the inclusions $i_\alpha : A_\alpha \hookrightarrow X$.

We now address the previously unanswered question about the homomorphism (12.1) from Lemma 11.18: what is its kernel?

We can make two immediate observations about this: first, for any group homomorphism $\Psi : G \to H$, $\ker \Psi$ is a normal subgroup of G. Recall that a subgroup $K \subset G$ is called **normal** if it is invariant under conjugation with arbitrary elements of G, i.e.

$$gkg^{-1} \in K \quad \text{for all } k \in K \text{ and } g \in G.$$

This condition is abbreviated by "$gKg^{-1} = K$". It is obviously satisfied if $K = \ker \Psi$ since $\Psi(k) = e$ implies $\Psi(gkg^{-1}) = \Psi(g)\Psi(k)\Psi(g^{-1}) = \Psi(g)e\Psi(g)^{-1} = e$. Recall further that for any subgroup $K \subset G$, the **quotient** G/K is defined as the set of all **left cosets** of K, meaning subsets of the form $gK := \{gh \mid h \in K\}$ for fixed elements $g \in G$. For arbitrary subgroups $K \subset G$, the quotient

G/K does not have a natural group structure, but it does when K is a *normal* subgroup: indeed, the condition $gKg^{-1} = K$ gives rise to a well-defined product

$$(aK)(bK) := (ab)K \in G/K$$

since, as subsets of G, $aKbK = a(bKb^{-1})bK = abKK = abK$. In particular, any homomorphism $\Psi : G \to H$ between groups G and H gives rise to a normal subgroup $K := \ker \Psi \subset G$ and thus a quotient group G/K, such that Ψ determines a a well-defined map

$$G/\ker \Psi \to H : gK \mapsto \Psi(g),$$

meaning that the value $\Psi(g)$ of this map does not depend on the choice of element $g \in G$ representing the coset $gK \in G/K$. It is easy to check that this map is also a group homomorphism, in which case we say that Ψ **descends** to a homomorphism $G/K \to H$, and moreover, it is injective since $\Psi(g) = e$ means $g \in \ker \Psi = K$ and thus $gK = K = eK$, which is the identity element of G/K. It follows that the induced map $G/\ker \Psi \to H$ is an isomorphism whenever the original homomorphism Ψ is surjective. (A standard reference for these basic notions from group theory is [**Art91**].)

The second observation concerns certain specific elements that obviously belong to the kernel of the map (12.1). Consider the inclusions

$$j_{\alpha\beta} : A_\alpha \cap A_\beta \hookrightarrow A_\alpha$$

for each pair $\alpha, \beta \in J$, and recall that $i_\alpha : A_\alpha \hookrightarrow X$ denotes the inclusion of $A_\alpha \subset X$. Then the following diagram commutes,

$$\begin{tikzcd}
 & A_\alpha \arrow[dr, "i_\alpha"] & \\
A_\alpha \cap A_\beta \arrow[ur, "j_{\alpha\beta}"] \arrow[dr, "j_{\beta\alpha}"'] & & X \\
 & A_\beta \arrow[ur, "i_\beta"'] &
\end{tikzcd}$$

meaning $i_\alpha \circ j_{\alpha\beta} = i_\beta \circ j_{\beta\alpha}$, since both are just the inclusion of $A_\alpha \cap A_\beta$ into X. This trivial observation has a nontrivial consequence for the homomorphism Φ. Indeed, for any loop $p \overset{\gamma}{\rightsquigarrow} p$ in $A_\alpha \cap A_\beta$ representing a nontrivial element of $\pi_1(A_\alpha \cap A_\beta, p)$, the two elements $(j_{\alpha\beta})_*[\gamma] \in \pi_1(A_\alpha, p)$ and $(j_{\beta\alpha})_*[\gamma] \in \pi_1(A_\beta, p)$ belong to distinct subgroups in the free product $*_{\gamma \in J}\, \pi_1(A_\gamma, p)$, yet clearly

$$(i_\alpha)_*(j_{\alpha\beta})_*[\gamma] = (i_\beta)_*(j_{\beta\alpha})_*[\gamma] \in \pi_1(X, p)$$

since $i_\alpha \circ j_{\alpha\beta} = i_\beta \circ j_{\beta\alpha}$. It follows that $\Phi((j_{\alpha\beta})_*[\gamma]) = \Phi((j_{\beta\alpha})_*[\gamma])$, hence $\ker \Phi$ must contain the reduced word formed by the two letters $(j_{\alpha\beta})_*[\gamma] \in \pi_1(A_\alpha, p)$ and $(j_{\beta\alpha})_*[\gamma]^{-1} \in \pi_1(A_\beta, p)$:

$$(j_{\alpha\beta})_*[\gamma](j_{\beta\alpha})_*[\gamma]^{-1} \in \ker \Phi.$$

Combining this with the first observation, $\ker \Phi$ must contain the smallest normal subgroup of $*_{\gamma \in J}\, \pi_1(A_\gamma, p)$ that contains all elements of this form.

DEFINITION 12.3. For any group G and subset $S \subset G$, we denote by

$$\langle S \rangle \subset G$$

the smallest subgroup of G that contains S, i.e. $\langle S \rangle$ is the set of all products of elements $g \in S$ and their inverses g^{-1}. Similarly,

$$\langle S \rangle_N \subset G$$

denotes the smallest *normal* subgroup of G that contains S. Concretely, this means $\langle S\rangle_N$ is the set of all conjugates of products of elements of S and their inverses.

We are now in a position to state the complete version of the Seifert-van Kampen theorem. The first half of the statement is just a repeat of Lemma 11.18, which we have proved already. The second half tells us what $\ker\Phi$ is, and thus gives a formula for $\pi_1(X,p)$.

THEOREM 12.4 (Seifert-van Kampen). *Suppose* $X = \bigcup_{\alpha\in J} A_\alpha$ *for a collection of open and path-connected subsets* $\{A_\alpha \subset X\}_{\alpha\in J}$ *with nonempty intersection, denote by* $i_\alpha : A_\alpha \hookrightarrow X$ *and* $j_{\alpha\beta} : A_\alpha \cap A_\beta \hookrightarrow A_\alpha$ *the inclusion maps for* $\alpha,\beta\in J$*, and fix* $p \in \bigcap_{\alpha\in J} A_\alpha$*.*

(1) *If* $A_\alpha \cap A_\beta$ *is path-connected for every pair* $\alpha,\beta\in J$*, then the natural homomorphism*

$$\Phi : \mathop{\ast}_{\alpha\in J} \pi_1(A_\alpha, p) \to \pi_1(X,p)$$

induced by the homomorphisms $(i_\alpha)_* : \pi_1(A_\alpha,p) \to \pi_1(X,p)$ *is surjective.*

(2) *If additionally* $A_\alpha\cap A_\beta\cap A_\gamma$ *is path-connected for every triple* $\alpha,\beta,\gamma\in J$*, then*

$$\ker\Phi = \left\langle \left\{ (j_{\alpha\beta})_*[\gamma](j_{\beta\alpha})_*[\gamma]^{-1} \;\middle|\; \alpha,\beta\in J,\ [\gamma]\in\pi_1(A_\alpha\cap A_\beta,p)\right\}\right\rangle_N .$$

In particular, Φ *then descends to an isomorphism*

$$\mathop{\ast}_{\alpha\in J} \pi_1(A_\alpha,p)\Big/ \ker\Phi \xrightarrow{\cong} \pi_1(X,p).$$

REMARK 12.5. In most applications, we will consider coverings of X by only two subsets $X = A\cup B$, and the condition on triple intersections in the second half of the statement then merely demands that $A\cap B$ be path-connected, which we already needed for the first half. (One can take the third subset in that condition to be either A or B; we never said that α, β and γ need to be distinct!)

I will give you the remaining part of the proof of this theorem in the next lecture. Let's now discuss some simple applications.

EXAMPLE 12.6. Consider the figure-eight $S^1\vee S^1$ with its natural base point $p\in S^1\vee S^1$, i.e. $S^1\vee S^1$ is the union of two circles $A,B\subset S^1\vee S^1$ with $A\cap B=\{p\}$. These are not open subsets, but since a neighborhood of p in $S^1\vee S^1$ has a fairly simple structure, we can get away with the usual trick (cf. Examples 11.3 and 11.7) of replacing both with homotopy equivalent open neighborhoods: define $A'\subset S^1\vee S^1$ as a small open neighborhood of A and $B'\subset S^1\vee S^1$ as a small open neighborhood of B such that there exist deformation retractions of A' to A and B' to B. The inclusions $A\hookrightarrow A'$ and $B\hookrightarrow B'$ then induce isomorphisms $\mathbb{Z}\cong\pi_1(A,p)\xrightarrow{\cong}\pi_1(A',p)$ and $\mathbb{Z}\cong\pi_1(B,p)\xrightarrow{\cong}\pi_1(B',p)$. The intersection $A'\cap B'$ is now a pair of line segments with one intersection point at p, so it admits a deformation retraction to p and is thus contractible, implying $\pi_1(A'\cap B',p)=0$. This makes $\ker\Phi$ in Theorem 12.4 trivial, hence the map

$$\pi_1(A,p) * \pi_1(B,p) \to \pi_1(S^1\vee S^1,p)$$

determined by the homomorphisms of $\pi_1(A,p)$ and $\pi_1(B,p)$ to $\pi_1(S^1\vee S^1,p)$ induced by the inclusions $A,B\hookrightarrow S^1\vee S^1$ is an isomorphism. To see more concretely what this group looks like, fix generators $\alpha\in\pi_1(A,p)\cong\mathbb{Z}$ and $\beta\in\pi_1(B,p)\cong\mathbb{Z}$, each of which can also be identified with elements of $\pi_1(S^1\vee S^1,p)$ via the inclusions of A and B into $S^1\vee S^1$. Then

$$\pi_1(S^1\vee S^1,p)\cong\mathbb{Z}*\mathbb{Z}=\{e,\alpha^p,\beta^q,\alpha^p\beta^q,\beta^p\alpha^q,\alpha^p\beta^q\alpha^r,\ldots \mid p,q,r,\ldots\in\mathbb{Z}\}.$$

These elements are easy to visualize: α and β are represented by loops that start and end at p and run once around the circles A or B respectively, so each element in the above list is a concatenation of finitely many repetitions of these two loops and their inverses. Notice that $\alpha\beta\neq\beta\alpha$, so $\pi_1(S^1\vee S^1)$ is our first example of a nonabelian fundamental group.

Example 12.7. Recall from Exercise 7.27 that for each $n \in \mathbb{N}$, one can identify S^n with the *one point compactification* of $\mathbb{R}^n$, a space defined by adjoining a single point called "∞" to $\mathbb{R}^n$:

$$S^n \cong \mathbb{R}^n \cup \{\infty\}.$$

This gives rise to an inclusion map $\mathbb{R}^n \overset{i}{\hookrightarrow} S^n$ with image $S^n \backslash \{\infty\}$. We claim that for any compact subset $K \subset \mathbb{R}^3$ such that $\mathbb{R}^3 \backslash K$ is path-connected, and any choice of base point $p \in \mathbb{R}^3 \backslash K$,

$$i_* : \pi_1(\mathbb{R}^3 \backslash K, p) \to \pi_1(S^3 \backslash K, p)$$

is an isomorphism. To see this, define the open subset $A := \mathbb{R}^3 \backslash K \subset S^3 \backslash K$, and choose $B_0 \subset S^3 \backslash K$ to be an open ball about ∞, i.e. a set of the form $(\mathbb{R}^3 \backslash \overline{B_R(0)}) \cup \{\infty\}$ where $\overline{B_R(0)} \subset \mathbb{R}^3$ is any closed ball large enough to contain K. Since p might not be contained in B_0 but $\mathbb{R}^3 \backslash K$ is path-connected, we can then define a larger set B by adjoining to B_0 the neighborhood in $\mathbb{R}^3 \backslash K$ of some path from a point in B_0 to p: this can be done so that both B_0 and B are homeomorphic to an open ball, so in particular they are contractible. The intersection $A \cap B$ is then $B \backslash \{\infty\}$ and is thus homoemorphic to $\mathbb{R}^3 \backslash \{0\}$ and homotopy equivalent to S^2, implying $\pi_1(A \cap B) = 0$. The Seifert-van Kampen theorem therefore gives an isomorphism $\pi_1(\mathbb{R}^3 \backslash K, p) * \pi_1(B, p) \to \pi_1(S^3 \backslash K, p)$, but $\pi_1(B, p)$ is the trivial group, so this proves the claim.

A frequently occuring special case of this example is when $K \subset \mathbb{R}^3$ is a knot, i.e. the image of an embedding $S^1 \hookrightarrow \mathbb{R}^3$. The fundamental group $\pi_1(\mathbb{R}^3 \backslash K)$ is then called the **knot group** of K, and the argument above shows that we are free to adjoin a point at infinity and thus replace the knot group with $\pi_1(S^3 \backslash K)$. This will be convenient for certain computations.

As in the previous lecture, we shall conclude this one by introducing some more terminology from combinatorial group theory in order to state a more usable variation on the Seifert-van Kampen theorem.

Definition 12.8. Given a set S, the **free group on** S is defined as

$$F_S := \mathop{\ast}_{\alpha \in S} \mathbb{Z},$$

or in other words, the set of all reduced words $a_1^{p_1} a_2^{p_2} \dots a_N^{p_N}$ for $N \geqslant 0$, $p_i \in \mathbb{Z}$ with $p_i \neq 0$, $a_i \in S$ and $a_i \neq a_{i+1}$ for every i, with the product defined by concatenation of words followed by reduction. The elements of S are called the **generators** of F_S.

Example 12.9. The computation in Example 12.6 gives $\pi_1(S^1 \vee S^1) \cong F_{\{\alpha,\beta\}} \cong \mathbb{Z} * \mathbb{Z}$, where the set generating $F_{\{\alpha,\beta\}}$ consists of the two loops α and β parametrizing the two circles that form $S^1 \vee S^1$.

Proposition 12.10. *For any set S, group G and map $\phi : S \to G$, there is a unique group homomorphism $\Phi : F_S \to G$ satisfying $\Phi(a) = \phi(a)$ for single-letter words $a \in F_S$ defined by elements $a \in S$.*

Proof. Writing elements of F_S in the form $a_1^{p_1} a_2^{p_2} \dots a_N^{p_N}$, there is clearly only one formula for $\Phi : F_S \to G$ that will match ϕ on single-letter words and also be a homomorphism, namely

$$\Phi(a_1^{p_1} \dots a_N^{p_N}) = \phi(a_1)^{p_1} \dots \phi(a_N)^{p_N}.$$

It is straightforward to check that this defines a homomorphism. □

Proposition 12.11. *Every group is isomorphic to a quotient of a free group by some normal subgroup.*

Proof. Pick any subset $S \subset G$ that generates G, e.g. one can choose $S := G$, though smaller subsets are usually also possible. Then the unique homomorphism $\Phi : F_S \to G$ sending each $g \in S \subset F_S$ to $g \in G$ is surjective, thus Φ descends to an isomorphism $F_S / \ker \Phi \to G$. □

Definition 12.12. Given a set S, a **relation** in S is defined to mean any equation of the form "$a = b$" where $a, b \in F_S$.

Definition 12.13. For any set S and a set R consisting of relations in S, we define the group

$$\{S \mid R\} := F_S \Big/ \langle R' \rangle_N$$

where R' is the set of all elements of the form $ab^{-1} \in F_S$ for relations "$a = b$" in R. The elements of S are called the **generators** of this group, and elements of R are its **relations**.

Let us pause a moment to interpret this definition. By a slight abuse of notation, we can write each element of $\{S \mid R\}$ as a reduced word w formed out of letters in S, with the understanding that w represents an equivalence class in the quotient $F_S/\langle R' \rangle_N$, thus it is possible to have $w = w'$ in $\{S \mid R\}$ even if w and w' are distinct elements of F_S. This will happen if and only if $w^{-1}w'$ belongs to the normal subgroup $\langle R' \rangle_N$, and in particular, it happens whenever "$w = w'$" is one of the relations in R. The relations are usually necesary because most groups are not free groups: while free groups are easy to describe (they depend only on their generators), most groups have more interesting structure than free groups, and this structure is encoded by relations. Proposition 12.11 implies that *every* group can be presented in this way, i.e. every group is isomorphic to $\{S \mid R\}$ for some set of generators S and relations R. Indeed, if $G = F_S/\ker \Phi$ for a set S and a surjective homomorphism $\Phi : F_S \to G$, then we can take S as the set of generators and define R to consist of all relations of the form "$a = b$" such that $ab^{-1} \in \ker \Phi$; the latter is equivalent to the condition $\Phi(a) = \Phi(b)$, so the relations tell us precisely when two products of generators give us the same element in G.

Definition 12.14. Given a group G, a **presentation** of G consists of a subset $S \subset G$ together with a set R of relations in S such that the unique homomorphism $F_S \to G$ matching the inclusion $S \hookrightarrow G$ on single-letter words descends to a group isomorphism

$$\{S \mid R\} \xrightarrow{\cong} G.$$

We say that G is **finitely presented** if it admits a presentation such that S and R are both finite sets.

Example 12.15. The group $\{a\} := \{a \mid \varnothing\}$ consisting of a single generator a with no relations is isomorphic to the free group $F_{\{a\}}$ on one element. The isomorphism $a^p \mapsto p$ identifies this with the integers $\mathbb{Z}$.

Example 12.16. The group $\{a, b \mid ab = ba\}$ has two generators and is abelian, so it is isomorphic to $\mathbb{Z}^2$. An explicit isomorphism is defined by $a^p b^q \mapsto (p, q)$. To see that this is an isomorphism, observe first that since $F_{\{a,b\}}$ is free, there exists a unique homomorphism $\Phi : F_{\{a,b\}} \to \mathbb{Z}^2$ with $\Phi(a) = (1, 0)$ and $\Phi(b) = (0, 1)$, and Φ is clearly surjective since it necesarily sends $a^p b^q$ to (p, q). Since $\mathbb{Z}^2$ is abelian, we also have

$$\Phi(ab(ba)^{-1}) = \Phi(aba^{-1}b^{-1}) = \Phi(a) + \Phi(b) - \Phi(a) - \Phi(b) = 0,$$

so $\ker \Phi$ contains $ab(ba)^{-1}$ and therefore also contains the smallest normal subgroup containing $ab(ba)^{-1}$, which is the group $\langle R' \rangle_N$ appearing in the quotient $\{a, b \mid ab = ba\} = F_{\{a,b\}}/\langle R' \rangle_N$. This proves that Φ descends to a surjective homomorphism $\{a, b \mid ab = ba\} \to \mathbb{Z}^2$. Finally, observe that since $ab = ba$ in the quotient $\{a, b \mid ab = ba\}$, every reduced word in $F_{\{a,b\}}$ is equivalent in this quotient to a word of the form $a^p b^q$ for some $(p, q) \in \mathbb{Z}^2$, and $\Phi(a^p b^q)$ then vanishes if and only if $a^p b^q = e$, proving that Φ is also injective.

Example 12.17. The group $\{a \mid a^p = e\}$ is isomorphic to $\mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$, with an explicit isomorphism defined in terms of the unique homomorphism $F_{\{a\}} \to \mathbb{Z}_p$ that sends a to $[1]$.

Example 12.18. We will prove in Lecture 14 that for the trefoil knot $K \subset \mathbb{R}^3 \subset S^3$, (see Lecture 8), $\pi_1(S^3 \backslash K) \cong \{a, b \mid a^2 = b^3\}$, and Exercise 12.20 below proves that this group is not abelian. By contrast, we will also see that the unknot $K_0 \subset \mathbb{R}^3 \subset S^3$ has $\pi_1(S^3 \backslash K_0) \cong \mathbb{Z}$, which is abelian. This implies via Example 12.7 that $\pi_1(\mathbb{R}^3 \backslash K) \not\cong \pi_1(\mathbb{R}^3 \backslash K_0)$, so $\mathbb{R}^3 \backslash K$ and $\mathbb{R}^3 \backslash K_0$ are not homeomorphic, hence the trefoil cannot be deformed continuously to the unknot.

Note that for any given set of generators S and relations R, it is often possible to reduce these to smaller sets without changing the isomorphism class of the group that they define. For the relations in particular, it is easy to imagine multiple distinct choices of the subset $R' \subset F_S$ that will produce the same normal subgroup $\langle R' \rangle_N$. In general, it is a very hard problem to determine whether or not two groups described via generators and relations are isomorphic; in fact, it is known that there does not exist any algorithm to decide whether a given presentation defines the trivial group. Nonetheless, generators and relations provide a very convenient way to describe many simple groups that arise in practice, especially in the context of van Kampen's theorem. This is due to the following reformulation of Theorem 12.4 for the case of two open subsets when all fundamental groups are finitely presented.

Corollary 12.19 (Seifert-van Kampen for finitely-presented groups). *Suppose $X = A \cup B$ where $A, B \subset X$ are open and path-connected subsets such that $A \cap B$ is also path-connected, and $j_A : A \cap B \hookrightarrow A$ and $j_B : A \cap B \hookrightarrow B$ denote the inclusions. Suppose moreover that there exist finite presentations*

$$\pi_1(A) \cong \big\{\{a_i\} \bigm| \{R_j\}\big\}, \qquad \pi_1(B) \cong \big\{\{b_k\} \bigm| \{S_\ell\}\big\}, \qquad \pi_1(A \cap B) \cong \big\{\{c_p\} \bigm| \{T_q\}\big\},$$

with the indices i, j, k, ℓ, p, q each ranging over finite sets. Then

$$\pi_1(X) \cong \big\{\{a_i\} \cup \{b_k\} \bigm| \{R_j\} \cup \{S_\ell\} \cup \{(j_A)_* c_p = (j_B)_* c_p\}\big\}.$$

□

In other words, as generators for $\pi_1(X)$, one can take all generators of $\pi_1(A)$ together with all generators of $\pi_1(B)$. The relations must then include all of the relations among the generators of $\pi_1(A)$ and $\pi_1(B)$ separately, but there may be additional relations that mix the generators from $\pi_1(A)$ and $\pi_1(B)$: these extra relations set $(j_A)_* c_p \in \pi_1(A)$ equal to $(j_B)_* c_p \in \pi_1(B)$ for each of the generators c_p of $\pi_1(A \cap B)$. These extra relations are exactly what is needed to describe the normal subgroup $\ker \Phi$ in the statement of Theorem 12.4. The relations in $\pi_1(A \cap B)$ do not play any role.

Exercise 12.20. Let us prove that the finitely-presented group $G = \{x, y \mid x^2 = y^3\}$ mentioned in Example 12.18 is nonabelian.

(a) Denoting the identity element by e, consider the related group

$$H = \{x, y \mid x^2 = y^3,\ y^3 = e,\ xyxy = e\}.$$

Show that every element of H is equivalent to one of the six elements $e, x, y, y^2, xy, xy^2 \in H$. This proves that H has order at most six, though in theory it could be less, since some of those six elements might still be equivalent to each other. To prove that this is not the case, construct (by writing down a multiplication table) a nonabelian group H' of order six that is generated by two elements a, b satisfying the relations $a^2 = b^3 = e$ and $abab = e$. Show that there exists a surjective homomorphism $H \to H'$, which is therefore an isomorphism since $|H| \leqslant 6$.

Remark: You don't need this fact, but if you've seen some of the standard examples of finite groups before, you might in any case notice that H is isomorphic to the dihedral group (Diedergruppe) of order 6.

(b) Show that H is a quotient of G by some normal subgroup, and deduce that G is also nonabelian.

EXERCISE 12.21. Given a group G, the **commutator subgroup** $[G, G] \subset G$ is the subgroup generated by all elements of the form

$$[x, y] := xyx^{-1}y^{-1}$$

for $x, y \in G$.

(a) Show that $[G, G] \subset G$ is always a normal subgroup, and it is trivial if and only if G is abelian.
(b) The **abelianization** (Abelisierung) of G is defined as the quotient group $G/[G, G]$. Show that this group is always abelian, and it is equal to G if G is already abelian.[13]
(c) Given any two abelian groups G, H, find a natural isomorphism from the abelianization of the free product $G * H$ to the Cartesian product $G \times H$.
(d) Prove that the abelianization of $\{x, y \mid x^2 = y^3\}$ is isomorphic to $\mathbb{Z}$.
Hint: An isomorphism φ from the abelianization to $\mathbb{Z}$ will be determined by two integers, $\varphi(x)$ and $\varphi(y)$. If φ exists, how must these two integers be related to each other?

13. Proof of the Seifert-van Kampen theorem

We have put off the proof of the Seifert-van Kampen theorem long enough. Here again is the statement.

THEOREM 13.1 (Seifert-van Kampen). *Suppose $X = \bigcup_{\alpha \in J} A_\alpha$ for a collection of open and path-connected subsets $\{A_\alpha \subset X\}_{\alpha \in J}$, $i_\alpha : A_\alpha \hookrightarrow X$ and $j_{\alpha\beta} : A_\alpha \cap A_\beta \hookrightarrow A_\alpha$ denote the natural inclusion maps for $\alpha, \beta \in J$, and $p \in \bigcap_{\alpha \in J} A_\alpha$.*

(1) If $A_\alpha \cap A_\beta$ is path-connected for every pair $\alpha, \beta \in J$, then the unique homomorphism

$$\Phi : \mathop{\circledast}_{\alpha \in J} \pi_1(A_\alpha, p) \to \pi_1(X, p)$$

that restricts to each subgroup $\pi_1(A_\alpha, p) \subset \circledast_{\beta \in J} \pi_1(A_\beta, p)$ as $(i_\alpha)_*$ is surjective.

(2) If additionally $A_\alpha \cap A_\beta \cap A_\gamma$ is path-connected for every triple $\alpha, \beta, \gamma \in J$, then

$$\ker \Phi = \langle S \rangle_N,$$

meaning $\ker \Phi$ is the smallest normal subgroup containing the set

$$S := \left\{ (j_{\alpha\beta})_*[\gamma](j_{\beta\alpha})_*[\gamma]^{-1} \;\middle|\; \alpha, \beta \in J,\; [\gamma] \in \pi_1(A_\alpha \cap A_\beta, p) \right\}.$$

In particular, if we abbreviate $F := \circledast_{\alpha \in J} \pi_1(A_\alpha, p)$, then Φ descends to an isomorphism

$$F \big/ \langle S \rangle_N \to \pi_1(X, p).$$

PROOF. We proved the first statement already in Lecture 11, so assume the hypothesis of the second statement holds. As observed in the previous lecture, $\Phi((j_{\alpha\beta})_*\gamma) = \Phi((j_{\beta\alpha})_*\gamma)$ for every $\alpha, \beta \in J$ and $\gamma \in \pi_1(A_\alpha \cap A_\beta, p)$, thus $\ker \Phi$ clearly contains $\langle S \rangle_N$, and in particular, Φ descends to a surjective homomorphism $F/\langle S \rangle_N \to \pi_1(X, p)$. We need to show that this homomorphism is injective, or equivalently, that whenever $\Phi(w) = \Phi(w')$ for a pair of reduced words $w, w' \in F$, their equivalence classes in $F/\langle S \rangle_N$ must match.

---

[13] Note that if $G = \{S \mid R\}$ is a finitely-presented group with generators S and relations R, then its abelianization is $\{S \mid R'\}$ where R' is the union of R with all relations of the form "$ab = ba$" for $a, b \in S$.

Given a loop $p \overset{\gamma}{\rightsquigarrow} p$ in X, let us say that a *factorization of* γ is any finite sequence $\{(\gamma_i, \alpha_i)\}_{i=1}^N$ such that $\alpha_i \in J$ and $p \overset{\gamma_i}{\rightsquigarrow} p$ is a loop in A_{α_i} for each $i = 1, \dots, N$, and

$$\gamma \underset{h+}{\sim} \gamma_1 \cdot \ldots \cdot \gamma_N.$$

The first half of the theorem follows from the fact (proved in Lemma 11.2) that every γ has a factorization. Now observe that any factorization as described above determines a reduced word $w \in F$, defined as the reduction of the word $[\gamma_1] \dots [\gamma_N]$ with $[\gamma_i] \in \pi_1(A_{\alpha_i}, p)$ for $i = 1, \dots, N$, and this word satisfies $\Phi(w) = [\gamma]$. Conversely, every reduced word $w \in \Phi^{-1}([\gamma])$ can be realized as a factorization of γ by choosing specific loops to represent the letters in w. The theorem will then follow if we can show that any two factorizations of γ can be related to each other by a finite sequence of the following operations and their inverses:

(A) Given two adjacent loops γ_i and γ_{i+1} such that $\alpha_i = \alpha_{i+1}$, replace them with their concatenation $p \overset{\gamma_i \cdot \gamma_{i+1}}{\rightsquigarrow} p$. (This does not change the corresponding reduced word in F, as it just implements a step in the reduction of an unreduced word.)

(B) Replace some γ_i with any loop γ_i' that is homotopic (with fixed end points) in A_{α_i}. (This also does not change the corresponding reduced word in F; in fact it doesn't even change the unreduced word from which it is derived.)

(C) Given a loop γ_i that lies in $A_{\alpha_i} \cap A_\beta$ for some $\beta \in J$, replace α_i with β. (In the corresponding reduced word in F, this replaces a letter of the form $(j_{\alpha_i\beta})_*[\gamma_i] \in \pi_1(A_{\alpha_i}, p)$ with one of the form $(j_{\beta\alpha_i})_*[\gamma_i] \in \pi_1(A_\beta, p)$, thus it changes the word but does not change its equivalence class in $F/\langle S \rangle_N$.)

We now prove that any two factorizations $\{(\gamma_i, \alpha_i)\}_{i=1}^N$ and $\{(\gamma_i', \alpha_i')\}_{i=1}^{N'}$ of γ are related by these operations. By assumption $\gamma_1 \cdot \ldots \cdot \gamma_N \underset{h+}{\sim} \gamma_1' \cdot \ldots \cdot \gamma_{N'}'$, so after choosing suitable parametrizations of both of these concatenations on the unit interval I,[14] there exists a homotopy

$$H : I^2 \to X$$

with $H(0, \cdot) = \gamma_1 \cdot \ldots \cdot \gamma_N$, $H(1, \cdot) = \gamma_1' \cdot \ldots \cdot \gamma_N'$ and $H(s, 0) = H(s, 1) = p$ for all $s \in I$. Since I^2 is compact, one can find a number $\epsilon > 0$ such that for every $(s, t) \in I^2$,[15] the intersection of I^2 with the box

$$[s - 2\epsilon, s + 2\epsilon] \times [t - 2\epsilon, t + 2\epsilon] \subset \mathbb{R}^2$$

is contained in $H^{-1}(A_\alpha)$ for some $\alpha \in J$. For suitably small $\epsilon = 1/n$ with $n \in \mathbb{N}$, we can therefore break up I^2 into n^2 boxes of side length ϵ which are each contained in $H^{-1}(A_\alpha)$ for some $\alpha \in J$ (possibly a different α for each box), forming a grid in I^2. For each box in the diagram there may be multiple $\alpha \in J$ that satisfy this condition, but let us choose a specific one to associate to each box. (These choices are indicated by the three colors in Figure 3.) Notice that each vertex in the grid is contained in the intersection of $H^{-1}(A_\alpha)$ for each of the $\alpha \in J$ associated to boxes that it touches. We can now perturb this diagram slightly to fill I^2 with a collection of boxes of slightly varying sizes such that every vertex in the interior touches only three of them (see the right side of Figure 3). We can similarly assume after such a perturbation that the vertices in $\{s = 0\}$ and $\{s = 1\}$ never coincide with the starting or ending times of the loops γ_i, γ_i' in the concatenations

---

[14] Recall that concatenation of paths is associative up to homotopy, so the N-fold concatenation $\gamma_1 \cdot \ldots \cdot \gamma_N$ is not a uniquely determined path $I \to X$ if $N > 2$, but it is unique up to homotopy with fixed end points.

[15] I do not consider this statement completely obvious, but it is a not very difficult exercise in point-set topology, and since that portion of the course is now over, I would rather leave it as an exercise than give the details here. Here is a hint: if the claim is not true, one can find a sequence $(s_k, t_k) \in I^2$ such that for each k, the intersection of I^2 with the box of side length $1/k$ about (s_k, t_k) is not fully contained in any of the subsets $H^{-1}(A_\alpha)$. This sequence has a convergent subsequence. What can you say about its limit?

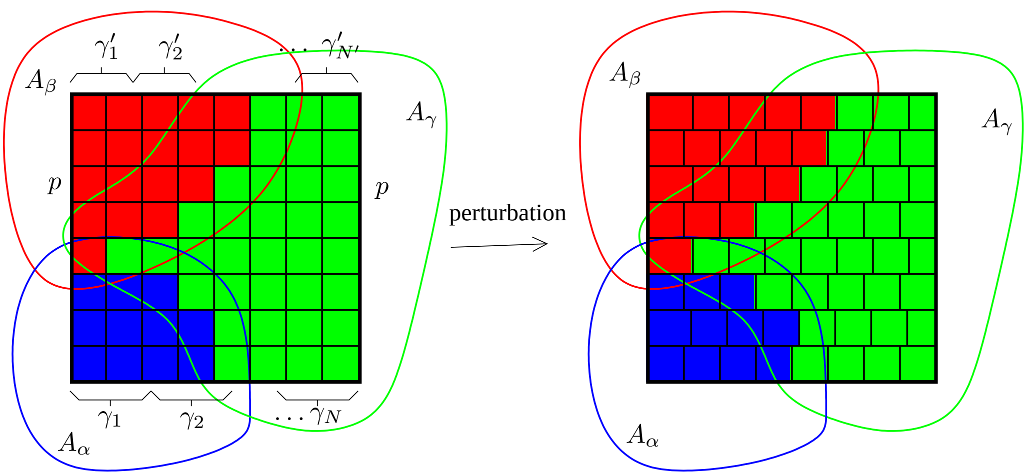


FIGURE 3. A grid on the domain of the homotopy $H : I^2 \to X$ between two factorizations $\gamma_1 \cdot \ldots \cdot \gamma_N$ and $\gamma_1' \cdot \ldots \cdot \gamma_{N'}'$ of a loop $p \overset{\gamma}{\rightsquigarrow} p$ in X. In this example, there are three open sets $A_\alpha, A_\beta, A_\gamma \subset X$, and colors are used to indicate that each of the small boxes filling I^2 has image lying in (at least) one of these subsets. In the perturbed picture at the right, every vertex in the interior touches exactly three boxes.

$\gamma_1 \cdot \ldots \cdot \gamma_N$ and $\gamma_1' \cdot \ldots \cdot \gamma_{N'}'$. Moreover, each vertex still lies in the same intersection of sets $H^{-1}(A_\alpha)$ as before, assuming the perturbation is sufficiently small.

Now suppose $(s,t) \in I^2$ is a vertex in the interior of the perturbed grid. Then (s,t) is on the boundary of exactly three boxes in the diagram, each of which belongs to one of the sets $H^{-1}(A_\alpha)$, $H^{-1}(A_\beta)$ and $H^{-1}(A_\gamma)$ for three associated elements $\alpha, \beta, \gamma \in J$ (they need not necessarily be distinct). If $(0,t)$ is a vertex with $t \notin \{0,1\}$, then it is on the boundary of exactly two boxes and thus lies in $H^{-1}(A_\alpha \cap A_\beta)$ for two associated elements $\alpha, \beta \in J$, but it also lies in $H^{-1}(A_\gamma)$ where $\gamma := \alpha_i$ is associated to the particular path γ_i whose domain as part of the concatenation $H(0,\cdot) = \gamma_1 \cdot \ldots \cdot \gamma_N$ contains $(0,t)$. For vertices $(1,t)$ with $t \notin \{0,1\}$, choose $A_\gamma := A_{\alpha_i'}$ similarly in terms of the concatenation $\gamma_1' \cdot \ldots \cdot \gamma_{N'}'$. In any of these cases, we have associated to each vertex (s,t) a path-connected set $A_\alpha \cap A_\beta \cap A_\gamma$ that contains $H(s,t)$, thus we can choose a path[16]

$$H(s,t) \overset{\delta_{(s,t)}}{\rightsquigarrow} p \quad \text{in} \quad A_\alpha \cap A_\beta \cap A_\gamma.$$

Since $H(s,t) = p$ for $t \in \{0,1\}$, this definition can be extended to vertices with $t \in \{0,1\}$ by defining $\delta_{(s,t)}$ as the trivial path. Now if E is any edge in the diagram, i.e. a side of one of the boxes, connecting two neighboring vertices (s_0,t_0) and (s_1,t_1), then we can identify E with the unit interval in order to regard $H|_E : E \to X$ as a path, and thus associate to E a loop

$$p \overset{\gamma_E}{\rightsquigarrow} p \quad \text{in} \quad A_\alpha \cap A_\beta, \qquad \gamma_E := \delta^{-1}_{(s_0,t_0)} \cdot H|_E \cdot \delta_{(s_1,t_1)},$$

where $\alpha, \beta \in J$ are the two (not necessarily distinct) elements associated to the boxes bordered by E.

---

[16] This is the specific step where we need the assumption that triple intersections are path-connected. If you're curious to see an example of the second half of the theorem failing without this assumption, I refer you to [**Hat02**, p. 44].

With these choices in place, any path through I^2 that follows a sequence of edges $E_1, \dots, E_k$ starting at some vertex in $(s_0, 0)$ and ending at a vertex $(s_1, 1)$ produces various factorizations of γ in the form $\{(\gamma_{E_i}, \beta_i)\}_{i=1}^k$. Here there is some freedom in the choices of $\beta_i \in J$: whenever a given edge E_i lies in $H^{-1}(A_\beta) \cap H^{-1}(A_\gamma)$, we can choose β_i to be either β or γ and thus produce two valid factorizations, which are related to each other by operation (C) in the list above.

We can now describe a procedure to modify the factorization $\{(\gamma_i, \alpha_i)\}_{i=1}^N$ to $\{(\gamma'_i, \alpha'_i)\}_{i=1}^{N'}$. We show first that $\{(\gamma_i, \alpha_i)\}_{i=1}^N$ is equivalent via our three operations to the factorization corresponding to the sequence of edges in $\{s = 0\}$ moving from $t = 0$ to $t = 1$. This is not so obvious because, although $H(0, \cdot)$ is a parametrization of the concatenated path $\gamma_1 \cdot \ldots \cdot \gamma_N$, the times that mark the boundaries between one path and the next in this concatenation need not have anything to do with the vertices of our chosen grid. Instead, our perturbation of the grid ensured that each γ_i in the concatenation hits vertices only in the interior of its domain, not at starting or end points. Denote by $(0, t_1), \dots, (0, t_{m-1})$ the particular grid vertices in the domain of γ_i, thus splitting up γ_i into a concatenation of paths $\gamma_i = \gamma_i^1 \cdot \ldots \cdot \gamma_i^m$ which have these vertices as starting and/or end points. Then

$$\gamma_i \underset{h+}{\sim} (\gamma_i^1 \cdot \delta_{(0,t_1)}) \cdot (\delta_{(0,t_1)}^{-1} \cdot \gamma_i^2 \cdot \delta_{(0,t_2)}) \cdot \ldots \cdot (\delta_{(0,t_{m-1})}^{-1} \cdot \gamma_i^m) \quad \text{ in } A_{\alpha_i}.$$

We can now apply operations (B) and (A) in that order to replace γ_i with the sequence of loops of the form $\delta_{(0,t_{j-1})}^{-1} \cdot \gamma_i^j \cdot \delta_{(0,t_j)}$ in A_{α_i} as indicated above. The result is a new factorization that has more loops in the sequence, but the resulting concatenation is broken up along points that include all vertices in $\{s = 0\}$. It is also broken along more points, corresponding to the pieces of the original concatenation $\gamma_1 \cdot \ldots \cdot \gamma_N$, but after applying operation (C) if necessary, we can now apply operation (A) to combine all adjacent loops whose domains belong to the same edge. The result is precisely the factorization corresponding to the sequence of edges in $\{s = 0\}$. The same procedure can be used to modify $\{(\gamma'_i, \alpha'_i)\}_{i=1}^{N'}$ to the factorization corresponding to the sequence of edges in $\{s = 1\}$.

To finish, we need to show that the factorization given by the edges in $\{s = 0\}$ can be transformed into the corresponding factorization at $\{s = 1\}$ by applying our three operations. The core of the idea for this is shown in Figure 4, where the purple curves show two sequences of edges which represent two factorizations. In this case the difference between one path and the other consists only of replacing two edges on adjacent sides of a particular box $Q \subset I^2$ with their two opposite sides, and we can change from one to the other as follows. First, if the box Q is in $H^{-1}(A_\alpha)$, apply the operation (C) to both factorizations until all the loops corresponding to sides of Q are regarded as loops in A_α. Having done this, both factorizations now contain two consecutive loops in A_α that correspond to two sides of Q, so we can apply the operation (A) to concatenate each of these pairs, reducing two loops to one distinguished loop through A_α in each factorization. Those two distinguished loops are also homotopic in A_α, as one can see by choosing a homotopy of paths through the square Q that connects two adjacent sides to their two opposite sides (Figure 4, right). This therefore applies the operation (B) to change one factorization to the other.

We note finally that for any sequence of edges that includes edges in $\{t = 0\}$ or $\{t = 1\}$, those edges represent the constant path at the base point p, and since concatenation with constant paths produces homotopic paths, adding these edges or removing them from the diagram changes the factorization by a combination of operations (A) and (B). It now only remains to observe that the path of edges along $\{s = 0\}$ can always be modified to the path of edges along $\{s = 1\}$ by a finite sequence of the modifications just described.

□

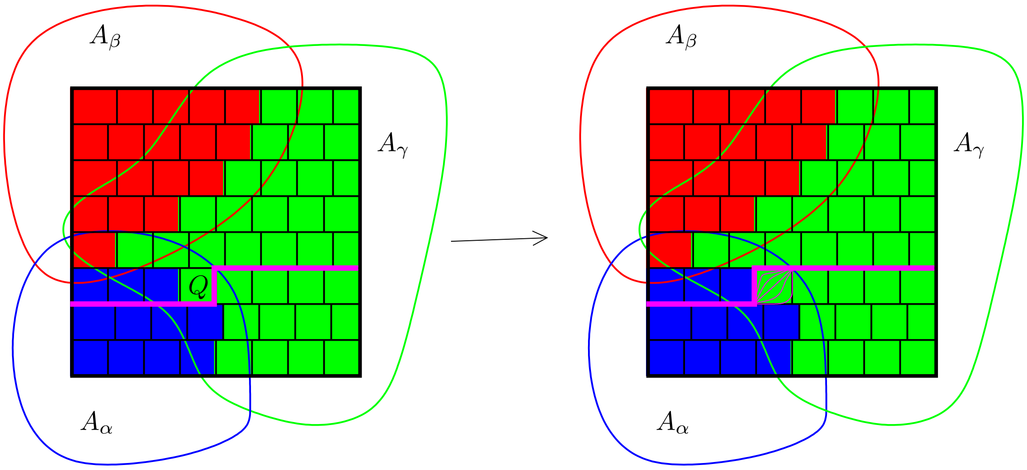


FIGURE 4. The magenta paths in both pictures are sequences of edges that define factorizations of γ, differing only at pairs of edges that surround a particular box Q. We can change one to the other by applying the three operations in our list.

EXERCISE 13.2. Recall that the wedge sum of two pointed spaces (X, x) and (Y, y) is defined as $X \vee Y = (X \amalg Y)/\sim$ where the equivalence relation identifies the two base points x and y. It is commonly said that whenever X and Y are both path-connected and are otherwise "reasonable" spaces, the formula

$$\pi_1(X \vee Y) \cong \pi_1(X) * \pi_1(Y) \tag{13.1}$$

holds. We saw for instance in Example 12.6 that this is true when X and Y are both circles. The goal of this problem is to understand slightly better what "reasonable" means in this context, and why such a condition is needed.

(a) Show by a direct argument (i.e. without trying to use Seifert-van Kampen) that if X and Y are both Hausdorff and simply connected, then $X \vee Y$ is simply connected.
Hint: Hausdorff implies that $X \backslash \{x\}$ and $Y \backslash \{y\}$ are both open subsets. Consider loops $\gamma : [0, 1] \to X \vee Y$ based at $[x] = [y]$ and decompose $[0, 1]$ into subintervals in which $\gamma(t)$ stays in either X or Y.

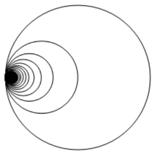
(b) Call a pointed space (X, x) *nice*[17] if x has an open neighborhood that admits a deformation retraction to x. Show that the formula (13.1) holds whenever (X, x) and (Y, y) are both nice, and more generally, the formula

$$\pi_1\left(\bigvee_{\alpha \in J} X_\alpha\right) \cong \mathop{\circledast}_{\alpha \in J} \pi_1(X_\alpha)$$

holds for any (possibly infinite) collection of nice pointed spaces $\{(X_\alpha, x_\alpha)\}_{\alpha \in J}$.

---

[17] Not a standardized term, I made it up.



(c) Here is an example of a space that is not "nice" in the sense of part (b): for each $n \in \mathbb{N}$, let $S^1_n \subset \mathbb{R}^2$ denote the circle of radius $1/n$ centered at $(1/n, 0)$. The union of all these circles is a space known informally as the *Hawaiian earring*

$$H := \bigcup_{n \in \mathbb{N}} S^1_n \subset \mathbb{R}^2.$$

As usual, we assign to H the subspace topology induced by the standard topology of $\mathbb{R}^2$. Show that in this space, the point $(0,0)$ does not have any simply connected open neighborhood.

(d) It is tempting to liken the Hawaiian earring H to the infinite wedge sum of circles $X := \bigvee_{n=1}^{\infty} S^1$, defined as above by choosing a base point in each copy of the circle and then identifying all the base points in the infinite disjoint union $\coprod_{n=1}^{\infty} S^1$. Both are unions of infinite collections of circles that all intersect each other at one point. Show in fact that there exists a continuous map

$$f : X \to H$$

that is a bijection sending the natural base point of $\bigvee_n S^1$ to $(0,0) \in \bigcap_n S^1_n$, but that X (unlike H) is a "nice" space, hence $f : X \to H$ cannot be a homeomorphism.
Hint: Continuity of maps defined on wedge sums is easy to check—see Exercise 10.5.

(e) Show that there exists a surjective continuous map $S^1 \to H$, but continuous maps $S^1 \to X$ are never surjective.
Hint: In H, start at $(0,0)$ and traverse the largest circle first, then continue to smaller circles.

(f) Show that for any finite subset $J \subset \mathbb{N}$, there exists a retraction

$$r_J : H \to \bigcup_{n \in J} S^1_n \subset H,$$

and deduce from this that the map $f_* : \pi_1(X) \to \pi_1(H)$ is injective.
Hint: Unlike H, $\bigcup_{n \in J} S^1_n$ really is homeomorphic to a wedge sum of circles, the crucial detail in this case being that there are only finitely many.

(g) Writing $r_n := r_{\{n\}} : H \to S^1_n$ for each individual value of $n \in \mathbb{N}$, show that the homomorphism

$$\pi_1(H) \to \prod_{n \in \mathbb{N}} \pi_1(S^1_n) \cong \prod_{n \in \mathbb{N}} \mathbb{Z}$$

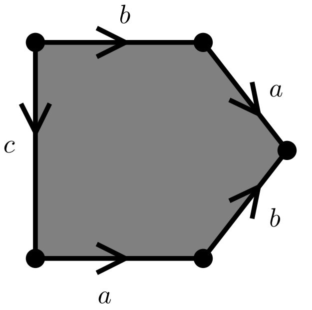
determined by the maps $(r_n)_* : \pi_1(H) \to \pi_1(S^1_n)$ is surjective, and deduce from this that $f_* : \pi_1(X) \to \pi_1(H)$ is not injective.
Remark: The direct product $\prod_{n \in \mathbb{N}} \mathbb{Z}$ of infinitely many groups (or in this case copies of the same group) is much larger than the direct sum $\bigoplus_{n \in \mathbb{N}} \mathbb{Z}$, and in fact, the standard "Cantor diagonal trick" that is typically used for proving the uncountability of $\mathbb{R}$ implies that $\prod_{n \in \mathbb{N}} \mathbb{Z}$ is likewise an uncountable set. It follows that $\pi_1(H)$ itself is uncountable, whereas $\pi_1(X) \cong \ast_{n \in \mathbb{N}} \mathbb{Z}$, being generated by countably many countable groups, is countable.

14. Surfaces and torus knots

We will discuss two applications of the Seifert-van Kampen theorem in this lecture: one to the study of surfaces, and the other to knots. Let's begin with surfaces.

Someday, when we talk about topological manifolds in this course (namely in Lecture 18), I will give you a precise mathematical definition of what the word "surface" means, but that day is not today. For now, we're just going to consider a class of specific examples that can be presented in a way that is convenient for computing their fundamental groups. A theorem we will discuss later in the semester implies that *all* compact surfaces can be presented in this way, but that is rather far from obvious.

We are going to consider pictures of polygons such as the following:

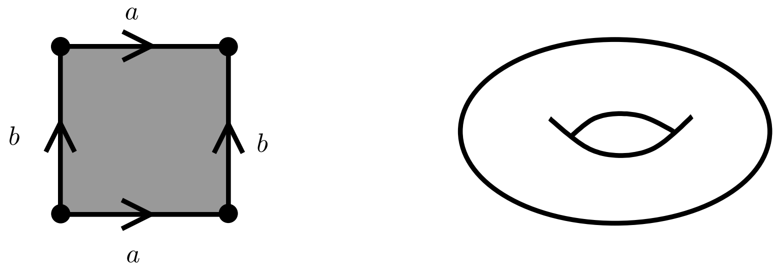


Suppose in general that $P \subset \mathbb{R}^2$ is a compact region bounded by some collection of N smooth curves that are arranged in a cyclic sequence with matching end points and do not intersect each other except at the matching end points. We will refer to these curves as *edges*, and label each of them with a letter a_i and an arrow. The letters $a_1, \ldots, a_N$ need not all be distinct. We then define a topological space

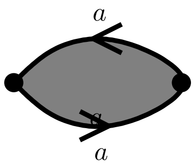
$$X := P/\sim,$$

where the equivalence relation is trivial on the interior of P but identifies all vertices with each other, thus collapsing the set of vertices to a single point, and it also identifies any pair of edges labeled by the same letter with each other via a homeomorphism that matches the directions of the arrows. (The exact choice of this homeomorphism will not matter.) In the picture above, this means the two edges labeled with "a" get identified, and so do the two edges labeled with "b". (By the time you've read to the end of this lecture, you should be able to form a fairly clear picture of this surface in your mind, but I suggest reading somewhat further before you try this.)

Example 14.1. Take P to be a square whose sides have two labels a and b such that opposite sides of the square have matching letters and arrows pointing in the same direction. You could then build a physical model of $X = P/\sim$ in two steps: take a square piece of paper and bend it until you can tape together the two opposite sides labeled a, producing a cylinder. The two boundary components of this cylinder are circles labeled b, so if you were doing this with a sufficiently stretchable material (paper is not stretchable enough), you could then bend the cylinder around and tape together its two circular boundary components. The result is what's depicted in the picture at the right, a space conventionally known as the **2-torus** (or just "the torus" for short) and denoted by $\mathbb{T}^2$. It is homeomorphic to the product $S^1 \times S^1$.



Example 14.2. If you relax your usual understanding of what a "polygon" is, you can also allow edges of the polygon to be curved as in the following example with only two edges:



The polygon itself is homeomorphic to the disk $\mathbb{D}^2$, but identifying the two edges via a homeomorphism matching the arrows means we identify each point on $\partial\mathbb{D}^2$ with its antipodal point. The result matches the second description of $\mathbb{RP}^2$ that we saw in the first lecture, see Example 1.2.

THEOREM 14.3. *Suppose $X = P/\sim$ is a space defined as described above by a polygon P with N edges labeled by (possibly repeated) letters $a_1, \ldots, a_N$, where we are listing them in the order in which they appear as the boundary is traversed once counterclockwise. Let G denote the set of all letters that appear in this list, and for each $i = 1, \ldots, N$, write $p_i = 1$ if the arrow at edge i points counterclockwise around the boundary and $p_i = -1$ otherwise. Then $\pi_1(X)$ is isomorphic to the group with generators G and exactly one relation $a_1^{p_1} \ldots a_N^{p_N} = e$:*

$$\pi_1(X) \cong \left\{ G \;\middle|\; a_1^{p_1} \ldots a_N^{p_N} = e \right\}.$$

PROOF. Let $P^1 := \partial P/\sim \subset X$. Since all vertices are identified to a point, P^1 is homeomorphic to a wedge sum of circles, one for each of the letters that appear as labels of edges, hence by an easy application of the Seifert-van Kampen theorem (cf. Exercise 13.2(b)),

$$\pi_1(P^1) \cong \pi_1(S^1) * \ldots * \pi_1(S^1) \cong \mathbb{Z} * \ldots * \mathbb{Z} = F_G,$$

the free group generated by the set G. Now decompose X into two open subsets A and B, where A is the interior of the polygon (not including its boundary) and B is an open neighborhood of P^1. We can arrange this so that $A \cap B$ is homeomorphic to an annulus $S^1 \times (-1, 1)$ occupying a neighborhood of ∂P in the interior of P, so for any choice of base point $p \in A \cap B$, $\pi_1(A \cap B, p) \cong \mathbb{Z}$ is generated by a loop that circles around parallel to ∂P. Since the neighborhood of ∂P admits a deformation retraction to ∂P, there is similarly a deformation retraction of B to P^1, giving $\pi_1(B, p) \cong \pi_1(P^1) = F_G$. Likewise, A is homeomorphic to an open disk, hence $\pi_1(A) = 0$. The Seifert-van Kampen theorem then idenifies $\pi_1(X, p)$ with a quotient of the free product $\pi_1(A, p) * \pi_1(B, p) \cong \pi_1(P^1) = F_G$, modulo the normal subgroup generated by the relation that if $j_A : A \cap B \hookrightarrow A$ and $j_B : A \cap B \hookrightarrow B$ denote the inclusion maps and $[\gamma] \in \pi_1(A \cap B, p) \cong \mathbb{Z}$ is a generator, then $(j_A)_*[\gamma] = (j_B)_*[\gamma]$. The left hand side of this equation is the trivial element since $\pi_1(A) = 0$. On the right hand side, we have the element of $\pi_1(B, p)$ represented by a loop $p \overset{\gamma}{\rightsquigarrow} p$ in the annulus $A \cap B$ that is parallel to the boundary of the polygon. Under the deformation retraction of $A \cap B$ to P^1, γ becomes the concatenated loop $a_1^{p_1} \ldots a_N^{p_N}$ defined by composing a traversal of ∂P with the quotient projection $\partial P \to P^1$, thus producing the relation $a_1^{p_1} \ldots a_N^{p_N} = e$. ☐

EXAMPLE 14.4. Applying the theorem to the torus in Example 14.1 gives

$$\pi_1(\mathbb{T}^2) \cong \{a, b \mid aba^{-1}b^{-1} = e\} = \{a, b \mid ab = ba\} \cong \mathbb{Z}^2.$$

Notice that this matches the result of applying Exercise 9.13(a), which gives $\pi_1(S^1 \times S^1) \cong \pi_1(S^1) \times \pi_1(S^1) \cong \mathbb{Z} \times \mathbb{Z}$.

EXAMPLE 14.5. For the picture of $\mathbb{RP}^2$ in Example 14.2, we obtain

$$\pi_1(\mathbb{RP}^2) \cong \{a \mid a^2 = e\} \cong \mathbb{Z}_2.$$

We already saw in Example 11.7 that $\pi_1(\mathbb{RP}^2)$ is generated by a single loop $\gamma : S^1 \to \mathbb{RP}^2$, the projection to $\mathbb{RP}^2 = S^2/\sim$ of a path that goes halfway around the equator of the sphere from one point to its antipodal point. We have now shown that $[\gamma]$ really is a nontrivial element of $\pi_1(\mathbb{RP}^2)$, but its square is trivial. The latter was also observed in Example 11.7, where it followed essentially from the fact that S^2 is simply connected: the concatenation of γ with itself is the projection to $\mathbb{RP}^2$ of a path that goes *all the way* around the equator in S^2, i.e. it is a loop, and can then be filled in with a map $\mathbb{D}^2 \to S^2$ since $\pi_1(S^2) = 0$. Composing the map $\mathbb{D}^2 \to S^2$ with the projection $S^2 \to \mathbb{RP}^2$ then contracts the loop γ^2 in $\mathbb{RP}^2$. However, we could not have deduced so easily from our knowledge of S^2 the fact that γ itself is *not* a contractible loop in $\mathbb{RP}^2$; that required the full strength of the Seifert-van Kampen theorem.

In Lecture 1, I drew you some pictures of topological spaces that I called "surfaces of genus g" for various values of a nonnegative integer g. I will now give you a precise definition of this space which, unfortunately, looks completely different from the original pictures, but we will soon see that it is equivalent.

DEFINITION 14.6. For any integer $g \geqslant 0$, the **closed orientable surface** Σ_g of **genus** (*Geschlecht*) g is defined to be S^2 if $g = 0$, and otherwise $\Sigma_g := P/\sim$ where P is a polygon with $4g$ edges labeled by $2g$ distinct letters $\{a_i, b_i\}_{i=1}^g$ in the order

$$a_1, b_1, a_1, b_1, a_2, b_2, a_2, b_2, \ldots, a_g, b_g, a_g, b_g,$$

such that the arrows point counterclockwise on the first instance of each letter in this sequence and clockwise on the second instance.

Once you've fully digested this definition, you may recognize that Σ_1 is defined by the square in Example 14.1, i.e. it is the torus $\mathbb{T}^2$. The diagram for Σ_2 is shown at the bottom of Figure 5. The projective plane $\mathbb{RP}^2$ is not an "orientable" surface, so it is not Σ_g for any g, though it is sometimes called a "non-orientable surface of genus 1". This terminology will make more sense when we later discuss the classification of surfaces.

In order to understand what Σ_g has to do with pictures we've seen before, we consider an operation on surfaces called the *connected sum*. It can be defined on any pair of surfaces Σ and Σ', or more generally, on any pair of n-dimensional topological manifolds, though for now we will consider only the case $n = 2$. Since I haven't yet actually given you precise definitions of the terms "surface" and "topological manifold," for now you should just assume Σ and Σ' come from the list of specific examples $\Sigma_0 = S^2$, $\Sigma_1 = \mathbb{T}^2$, $\Sigma_2, \Sigma_3, \ldots$ defined above.

Given a pair of inclusions $\mathbb{D}^2 \hookrightarrow \Sigma$ and $\mathbb{D}^2 \hookrightarrow \Sigma'$, the **connected sum** (*zusammenhängende Summe*) of Σ and Σ' is defined as the space

$$\Sigma \# \Sigma' := \left(\Sigma \backslash \mathring{\mathbb{D}}^2\right) \cup_{S^1} \left(\Sigma' \backslash \mathring{\mathbb{D}}^2\right).$$

The result of this operation is not hard to visualize in many concrete examples, see e.g. Figure 6.

More generally, for topological n-manifolds M and M', one defines the connected sum $M \# M'$ by choosing inclusions of $\mathbb{D}^n$ into M and M', then removing the interiors of these disks and gluing together $M \backslash \mathring{\mathbb{D}}^n$ and $M' \backslash \mathring{\mathbb{D}}^n$ along $S^{n-1} = \partial \mathbb{D}^n$. The notation $M \# M'$ obscures the fact that the definition of the connected sum depends explicitly on choices of inclusions of $\mathbb{D}^n$ into both spaces, and it is not entirely true in general that $M \# M'$ up to homeomorphism is independent of this choice. It is true however for surfaces:

LEMMA 14.7 (slightly nontrivial). *Up to homeomorphism, the connected sum $\Sigma \# \Sigma'$ of two closed connected surfaces Σ and Σ' does not depend on the choices of inclusions $\mathbb{D}^2 \hookrightarrow \Sigma$ and $\mathbb{D}^2 \hookrightarrow \Sigma'$.*

SKETCH OF A PROOF. A complete proof of this would be too much of a digression and require more knowledge about the classification of surfaces than is presently safe to assume, but I can

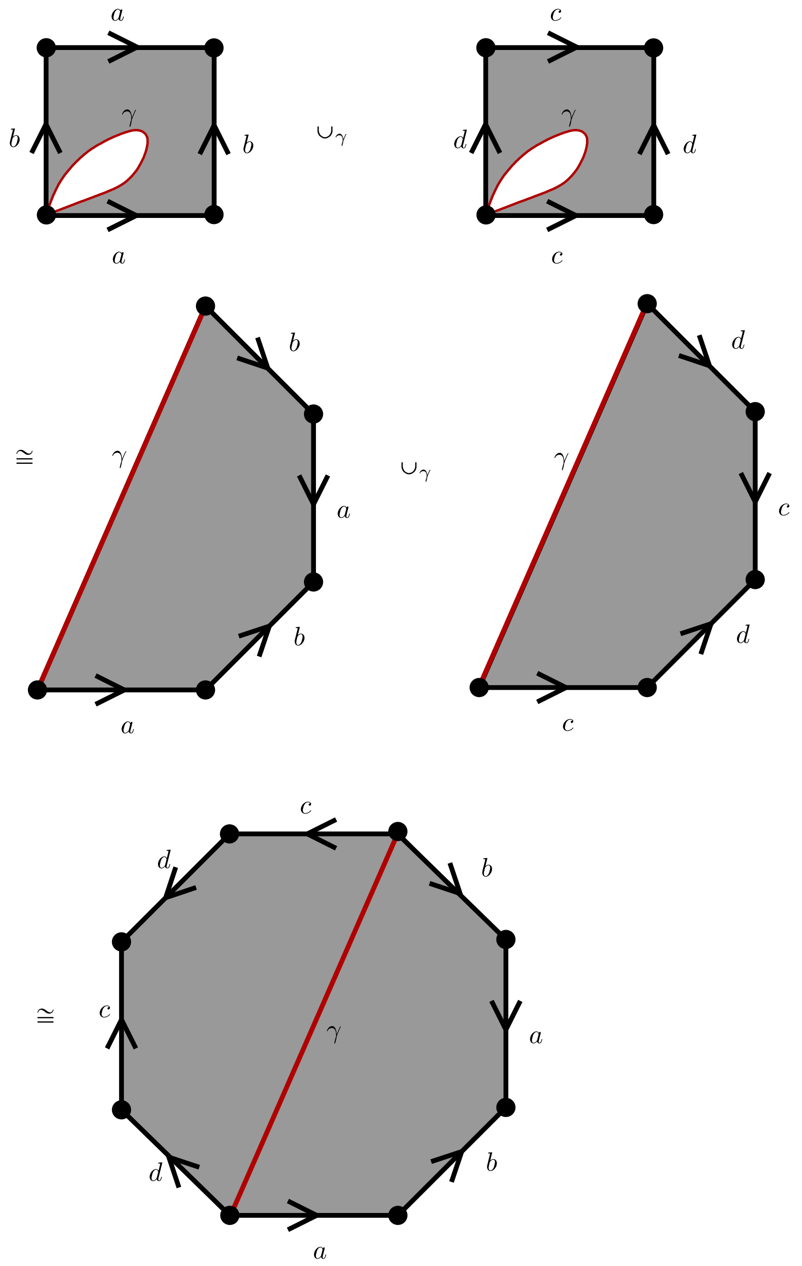


Figure 5. The connected sum $\mathbb{T}^2\#\mathbb{T}^2$ is formed by cutting holes $\mathbb{D}^2$ out of two copies of $\mathbb{T}^2$ along some loop γ, and then gluing together the two copies of $\mathbb{T}^2\backslash\mathbb{D}^2$. The result is Σ_2, the closed orientable surface of genus 2.

give the rough idea. The main thing you need to believe is that "up to orientation" (I'll come back to that detail in a moment), any inclusion $i_0 : \mathbb{D}^2 \hookrightarrow \Sigma$ can be deformed into any other inclusion $i_1 : \mathbb{D}^2 \hookrightarrow \Sigma$ through a continuous family of inclusions $i_t : \mathbb{D}^2 \hookrightarrow \Sigma$ for $t \in I$. You should imagine this roughly as follows: since $\mathbb{D}^2$ is homeomorphic via the obvious rescalings to the disk

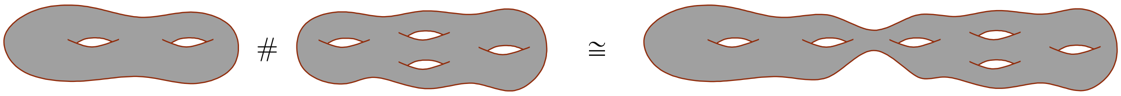


FIGURE 6. The connected sum of two surfaces is defined by cutting a hole out of each of them and gluing the rest together along the resulting boundary circle.

$\mathbb{D}^2_r$ of radius r for every $r > 0$, one can first deform i_0 and i_1 to inclusions whose images lie in arbitrarily small neighborhoods of two points $z_0, z_1 \in \Sigma$. Now since Σ is connected (and therefore also path-connected, as all topological manifolds are locally path-connected), we can choose a path γ from z_0 to z_1, and the idea is then to define i_t as a continuous family of inclusions $\mathbb{D}^2 \hookrightarrow \Sigma$ such that the image of i_t lies in an arbitrarily small neighborhood of $\gamma(t)$ for each t. You should be able to imagine concretely how to do this in the special case $\Sigma = \mathbb{R}^2$. That it can be done on arbitrary connected surfaces Σ depends on the fact that every point in Σ has a neighborhood homeomorphic to $\mathbb{R}^2$ (in other words, Σ is a topological 2-manifold).

Now for the detail that was brushed under the rug in the previous paragraph: even if $i_0, i_1 : \mathbb{D}^2 \hookrightarrow \Sigma$ are two inclusions that send 0 to the same point $z \in \Sigma$ and have images in an arbitrarily small neighborhood of z, it is not always true that i_0 can be deformed to i_1 through a continuous family of inclusions. For example, if we take $\Sigma = \mathbb{R}^2$, it is not true for the two inclusions $i_0, i_1 : \mathbb{D}^2 \hookrightarrow \mathbb{R}^2$ defined by $i_0(x, y) = (\epsilon x, \epsilon y)$ and $i_1(x, y) = (\epsilon x, -\epsilon y)$. In this example, both inclusions are defined as restrictions of injective linear maps $\mathbb{R}^2 \to \mathbb{R}^2$, but one has positive determinant and the other has negative determinant, so one cannot deform from one to the other through injective linear maps. One can use the technology of *local homology groups* (which we'll cover next semester) to remove the linearity from this argument and show that there also is no deformation from i_0 to i_1 through continuous inclusions. The issue here is one of *orientations*: i_0 is an orientation-preserving map, while i_1 is orientation-reversing. It turns out that two inclusions of $\mathbb{D}^2$ into $\mathbb{R}^2$ can be deformed to each other through inclusions if and only if they are either both orientation preserving or both orientation reversing. This obstruction sounds like bad news for our proof, but the situation is saved by the following corollary of the classification of surfaces: every closed orientable surface admits an orientation-reversing homeomorphism to itself. For example, if you picture the torus as the usual tube embedded in $\mathbb{R}^3$ and you embed it so that it is *symmetric* about some 2-dimensional coordinate plane, then the linear reflection through that plane restricts to a homeomorphism of $\mathbb{T}^2$ that is orientation reversing. Once we see what all the other closed orientable surfaces look like, it will be easy to see that one can do that with all of them. Actually, it is also not so hard to see this for the surfaces Σ_g defined as polygons: you just need to choose a sufficiently clever axis in the plane containing the polygon and reflect across it. Once this is understood, you realize that the orientation of your inclusion $\mathbb{D}^2 \hookrightarrow \Sigma$ does not really matter, as you can always replace it with an inclusion having the opposite orientation, and the picture you get in the end will be homeomorphic to the original.

With this detail out of the way, you just have to convince yourself that if you have a pair of continuous families of inclusions $i_t : \mathbb{D}^2 \hookrightarrow \Sigma$ and $j_t : \mathbb{D}^2 \hookrightarrow \Sigma'$ defined for $t \in [0, 1]$, then the resulting glued surfaces

$$\Sigma \#_t \Sigma' := \left(\Sigma \backslash i_t(\mathring{\mathbb{D}}^2)\right) \cup_{S^1} \left(\Sigma' \backslash j_t(\mathring{\mathbb{D}}^2)\right)$$

are homeomorphic for all t. It suffices in fact to prove that this is true just for t varying in an arbitrarily small interval $(t_0 - \epsilon, t_0 + \epsilon)$, since $[0, 1]$ is compact and can therefore be covered by finitely many such intervals. A homeomorphism $\Sigma \#_t \Sigma' \to \Sigma \#_s \Sigma'$ for $t \neq s$ is easy to define if we can first find a homeomorphism $\Sigma \to \Sigma$ that sends $i_t(z) \mapsto i_s(z)$ for every $z \in \mathbb{D}^2$ and similarly on Σ'. This is not hard to construct if t and s are sufficiently close. ☐

Now we are in a position to relate Σ_g with the more familiar pictures of surfaces.

THEOREM 14.8. *For any nonnegative integers g, h, $\Sigma_g \# \Sigma_h \cong \Sigma_{g+h}$. In particular, Σ_g is the connected sum of g copies of the torus:*

$$\Sigma_g \cong \underbrace{\mathbb{T}^2 \# \ldots \# \mathbb{T}^2}_{g}$$

PROOF. The result becomes obvious if one makes a sufficiently clever choice of hole to cut out of Σ_g and Σ_h, and Lemma 14.7 tells us that the resulting space up to homeomorphism is independent of this choice. The example of $g = h = 1$ is shown in Figure 5, and the same idea works (but is more effort to draw) for any values of g and h. ☐

Now that we know how to draw pretty pictures of the surfaces Σ_g, we can also observe that we have already proved something quite nontrivial about them: we have computed their fundamental groups!

COROLLARY 14.9 (of Theorem 14.3). *The closed orientable surface Σ_g of genus $g \geqslant 0$ has a fundamental group with $2g$ generators and one relation, namely*

$$\pi_1(\Sigma_g) \cong \left\{a_1, b_1, \ldots, a_g, b_g \;\middle|\; a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \ldots a_g b_g a_g^{-1} b_g^{-1} = e\right\}.$$

☐

Using the commutator notation from Exercise 12.21, the relation in Corollary 14.9 can be conveniently abbreviated as

$$\prod_{i=1}^{g} [a_i, b_i] = e.$$

EXERCISE 14.10. Show that the abelianization (cf. Exercise 12.21) of $\pi_1(\Sigma_g)$ is isomorphic to the additive group $\mathbb{Z}^{2g}$.
Hint: $\pi_1(\Sigma_g)$ is a particular quotient of the free group on $2g$ generators. Observe that the abelianization of that free group is identical to the abelianization of $\pi_1(\Sigma_g)$. (Why?)

By the classification of finitely generated abelian groups, $\mathbb{Z}^m$ and $\mathbb{Z}^n$ are never isomorphic unless $m = n$, so Exercise 14.10 implies that $\pi_1(\Sigma_g)$ and $\pi_1(\Sigma_h)$ are not isomorphic unless $g = h$. This completes the first step in the classification of closed surfaces:

COROLLARY 14.11. *For two nonnegative integers $g \neq h$, Σ_g and Σ_h are not homeomorphic.* ☐

EXERCISE 14.12. Assume X and Y are path-connected topological manifolds of dimension n.

(a) Use the Seifert-Van Kampen theorem to show that if $n \geqslant 3$, then $\pi_1(X \# Y) \cong \pi_1(X) * \pi_1(Y)$. Where does your proof fail in the cases $n = 1$ and $n = 2$?
(b) Show that the formula of part (a) is false in general for $n = 1, 2$.

EXERCISE 14.13. For integers $g, m \geqslant 0$, let $\Sigma_{g,m}$ denote the compact surface obtained by cutting m disjoint disk-shaped holes out of the closed orientable surface with genus g. (By this convention, $\Sigma_g = \Sigma_{g,0}$.) The boundary $\partial \Sigma_{g,m}$ is then a disjoint union of m circles, e.g. the case with $g = 1$ and $m = 3$ is shown in Figure 7.

(a) Show that $\pi_1(\Sigma_{g,1})$ is a free group with $2g$ generators, and if $g \geqslant 1$, then any simple closed curve parametrizing $\partial \Sigma_{g,1}$ represents a nontrivial element of $\pi_1(\Sigma_{g,1})$.[18]
Hint: Think of Σ_g as a polygon with some of its edges identified. If you cut a hole in the middle of the polygon, what remains admits a deformation retraction to the edges. Prove it with a picture.

---

[18] Terminology: one says in this case that $\partial \Sigma_{g,1}$ is **homotopically nontrivial** or **essential**, or equivalently, $\partial \Sigma_{g,1}$ is not **nullhomotopic**.

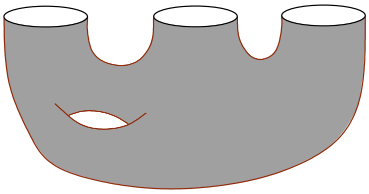
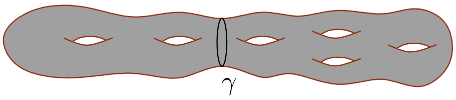


FIGURE 7. The surface $\Sigma_{1,3}$ as in Exercise 14.13.



(b) Assume γ is a simple closed curve separating Σ_g into two pieces homeomorphic to $\Sigma_{h,1}$ and $\Sigma_{k,1}$ for some $h, k \geqslant 0$. (The picture at the right shows an example with $h = 2$ and $k = 4$.) Show that the image of $[\gamma] \in \pi_1(\Sigma_g)$ under the natural projection to the abelianization of $\pi_1(\Sigma_g)$ is trivial.
Hint: What does γ look like in the polygonal picture from part (a)? What is it homotopic to?

(c) Prove that if $g \geqslant 2$ and G denotes the group $\{a_1, b_1, \dots, a_g, b_g \mid \prod_{i=1}^g [a_i, b_i] = e\}$, then for any proper subset $J \subset \{1, \dots, g\}$, $\prod_{i \in J}[a_i, b_i]$ is a nontrivial element of G.
Hint: Given $j \in J$ and $\ell \in \{1, \dots, g\} \backslash J$, there is a homomorphism $\Phi : F_{\{a_1, b_1, \dots, a_g, b_g\}} \to F_{\{x,y\}}$ that sends $a_j \mapsto x$, $b_j \mapsto y$, $a_\ell \mapsto y$, $b_\ell \mapsto x$ and maps all other generators to the identity. Show that Φ descends to the quotient G and maps $\prod_{i \in J}[a_i, b_i] \in G$ to something nontrivial.

(d) Deduce from part (c) that if $h > 0$ and $k > 0$, then the curve γ in part (b) represents a nontrivial element of $\pi_1(\Sigma_g)$.

(e) Generalize part (a): show that if $m \geqslant 1$, $\pi_1(\Sigma_{g,m})$ is a free group with $2g + m - 1$ generators.

Now let's talk about knots. Back in Lecture 8, I showed you two simple examples of knots $K \subset \mathbb{R}^3$: the *trefoil* and the *unknot*. I claimed that it is impossible to deform one of these knots into the other, and in fact that the complements of both knots in $\mathbb{R}^3$ are not homeomorphic. It is time to prove this.

We will consider both as special cases of a more general class of knots called *torus knots*. Fix the standard embedding of the torus

$$f : \mathbb{T}^2 = S^1 \times S^1 \hookrightarrow \mathbb{R}^3,$$

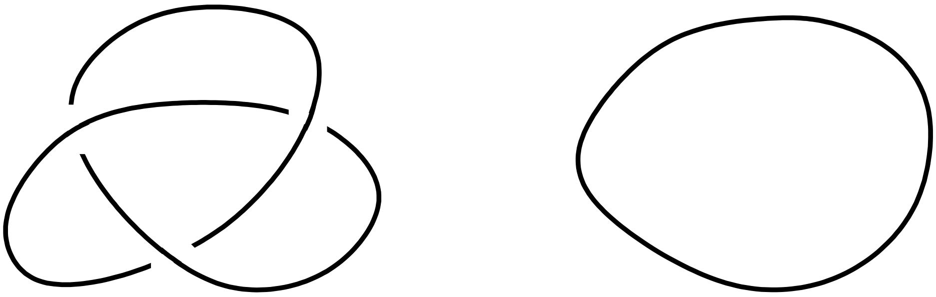
where by "standard," I mean the one that you usually picture when you imagine a torus embedded in $\mathbb{R}^3$ (see the surface bounding the grey region in Figure 9). Given any two relatively prime integers $p, q \in \mathbb{Z}$, the **(p, q)-torus knot** is defined by

$$K_{p,q} := \{ f(e^{pi\theta}, e^{qi\theta}) \mid \theta \in \mathbb{R} \} \subset \mathbb{R}^3.$$

In other words, $K_{p,q}$ is a knot lying on the image of the embedded torus $f(\mathbb{T}^2) \subset \mathbb{R}^3$, obtained from a loop that rotates p times around one of the dimensions of $\mathbb{T}^2 = S^1 \times S^1$ while rotating q times around the other. It is conventional to assume p and q are relatively prime, since the definition of $K_{p,q}$ above would not change if both p and q were multiplied by the same nonzero constant.

EXAMPLE 14.14. $K_{2,3}$ is the trefoil knot (Figure 8, left).

EXAMPLE 14.15. $K_{1,0}$ is the unknot (Figure 8, right).



$K_{2,3} \subset \mathbb{R}^3$ $K_{1,0} \subset \mathbb{R}^3$

FIGURE 8. The trefoil knot $K_{2,3}$ and unknot $K_{1,0}$.

The **knot group** of a knot $K \subset \mathbb{R}^3$ is defined as the fundamental group of the so-called *knot complement*, $\pi_1(\mathbb{R}^3 \backslash K)$. We saw in Example 12.7 that the natural inclusion $\mathbb{R}^3 \hookrightarrow S^3$ defined by identifying S^3 with the one-point compactification $\mathbb{R}^3 \cup \{\infty\}$ induces an isomorphism of $\pi_1(\mathbb{R}^3 \backslash K)$ to $\pi_1(S^3 \backslash K)$, thus in order to compute knot groups, we may as well regard the knot $K \subset \mathbb{R}^3$ as a subset of the slightly larger but *compact* space S^3 and compute $\pi_1(S^3 \backslash K)$. We shall now answer the question: given relatively prime integers p and q, what is $\pi_1(S^3 \backslash K_{p,q})$?

Here is a useful trick for picturing S^3. By definition, $S^3 = \partial \mathbb{D}^4$, but notice that $\mathbb{D}^4$ is also homeomorphic to the "box" $\mathbb{D}^2 \times \mathbb{D}^2$, whose boundary consists of the two pieces $\partial \mathbb{D}^2 \times \mathbb{D}^2$ and $\mathbb{D}^2 \times \partial \mathbb{D}^2$, intersecting each other along $\partial \mathbb{D}^2 \times \partial \mathbb{D}^2$. The latter is a copy of $\mathbb{T}^2$, and the pieces $S^1 \times \mathbb{D}^2$ and $\mathbb{D}^2 \times S^1$ are called **solid tori** since we usually picture them as the region in $\mathbb{R}^3$ bounded by the standard embedding of the torus. The homeomorphism $\mathbb{D}^4 \cong \mathbb{D}^2 \times \mathbb{D}^2$ thus allows us to identify S^3 with the space constructed by gluing together these two solid tori along the obvious identification of their boundaries:

$$S^3 \cong (S^1 \times \mathbb{D}^2) \cup_{\mathbb{T}^2} (\mathbb{D}^2 \times S^1).$$

A picture of this decomposition is shown in Figure 9. Here the 2-torus along which the two solid tori are glued together is depicted as the standard embedding of $\mathbb{T}^2$ in $\mathbb{R}^3$, so this is where we will assume $K_{p,q}$ lies. The region bounded by this torus is $S^1 \times \mathbb{D}^2$, shown in the picture as an S^1-parametrized family of disks $\mathbb{D}^2$. It requires a bit more imagination to recognize $\mathbb{D}^2 \times S^1$ in the picture: instead of a family of disks, we have drawn it as a $\mathbb{D}^2$-parametrized family of circles, where it is important to understand that one of those circles passes through $\infty \in S^3$ and thus looks like a line instead of a circle in the picture. This picture will now serve as the basis for a Seifert-van Kampen decomposition of $S^3 \backslash K_{p,q}$ into two open subsets. They will be defined as open neighborhoods of the two subsets

$$A_0 := (S^1 \times \mathbb{D}^2) \backslash K_{p,q}, \qquad B_0 := (\mathbb{D}^2 \times S^1) \backslash K_{p,q}.$$

In order to define suitable neighborhoods, let us identify a neighborhood of $f(\mathbb{T}^2)$ in $\mathbb{R}^3$ with $(-1,1) \times \mathbb{T}^2$ such that $f(\mathbb{T}^2)$ becomes $\{0\} \times \mathbb{T}^2 \subset \mathbb{R}^3$. We then define

$$A := \left(S^1 \times \mathring{\mathbb{D}}^2\right) \cup \left((-1,1) \times (\mathbb{T}^2 \backslash f^{-1}(K_{p,q}))\right),$$

and

$$B := \left(\mathring{\mathbb{D}}^2 \times S^1\right) \cup \left((-1,1) \times (\mathbb{T}^2 \backslash f^{-1}(K_{p,q}))\right).$$

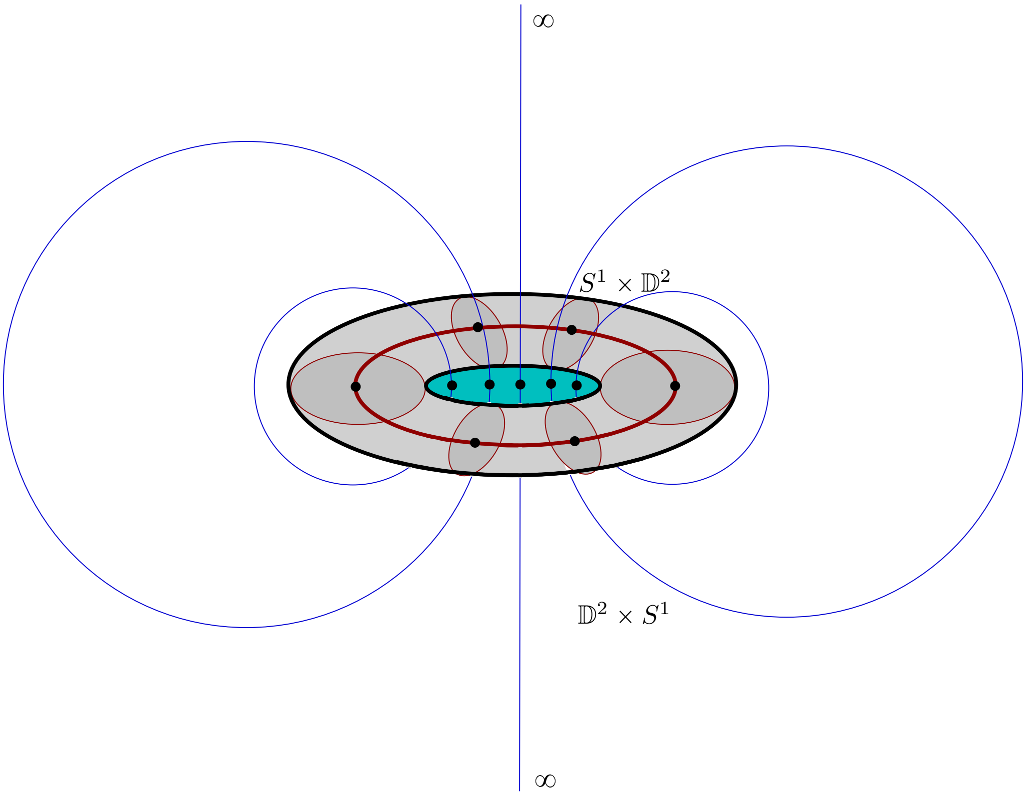


FIGURE 9. The sphere $S^3 = \mathbb{R}^3 \cup \{\infty\}$ decomposed as a union of two solid tori whose common boundary is the "standard" embedding of $\mathbb{T}^2$ in $\mathbb{R}^3$: $S^3 \cong \partial(\mathbb{D}^2 \times \mathbb{D}^2) = (S^1 \times \mathbb{D}^2) \cup_{\mathbb{T}^2} (\mathbb{D}^2 \times S^1)$. The vertical blue line passing through the middle is actually a circle in S^3 passing through the point at ∞.

By contracting the interval $(-1, 1)$, we can define a deformation retraction of A to A_0 and then retract further by contractng the disk $\mathbb{D}^2$ to its center, eventually producing a deformation retraction of A to the circle $S^1 \times \{0\}$ at the center of the inner solid torus—this is the red circle in Figure 9 that passes through the center of each disk. In an analogous way, there is a deformation retraction of B to the center $\{0\} \times S^1$ of the outer solid torus, which is the blue line through ∞ in the picture, though you might prefer to perturb this to one of the parallel circles $\{z\} \times S^1 \subset \mathbb{D}^2 \times S^1$ for $z \neq 0$, since these actually look like circles in the picture. We can now regard $\pi_1(A)$ and $\pi_1(B)$ as separate copies of the integers whose generators we shall call a and b respectively,

$$\pi_1(A) \cong \{a \mid \varnothing\}, \qquad \pi_1(B) \cong \{b \mid \varnothing\}.$$

The intersection is

$$A \cap B = (-1, 1) \times \left(\mathbb{T}^2 \backslash f^{-1}(K_{p,q})\right) \underset{h.e.}{\simeq} \mathbb{T}^2 \backslash f^{-1}(K_{p,q}) \underset{h.e.}{\simeq} S^1.$$

That last homotopy equivalence deserves an explanation: if you draw $\mathbb{T}^2$ as a square with its sides identified, then $f^{-1}(K_{p,q})$ looks like a straight line that periodically exits one side of the square and reappears at the opposite side. Now draw another straight path parallel to this one (I recommend using a different color), and you will easily see that after removing $f^{-1}(K_{p,q})$ from $\mathbb{T}^2$, what remains admits a deformation retraction to the parallel path, which is an embedded copy of S^1. We will call the generator of its fundamental group c,

$$\pi_1(A \cap B) \cong \{c \mid \varnothing\}.$$

According to the Seifert-van Kampen theorem (in particular Corollary 12.19, the version for finitely-presented groups), we can now write

$$\pi_1(S^3 \backslash K_{p,q}) \cong \left\{a, b \,\middle|\, (j_A)_* c = (j_B)_* c\right\},$$

where j_A and j_B denote the inclusions of $A \cap B$ into A and B respectively. To interpret this properly, we should choose a base point in $A \cap B$ and picture a, b and c as represented by specific loops through this base point, so without loss of generality, a is a loop near the boundary $\mathbb{T}^2$ of $S^1 \times \mathbb{D}^2$ that wraps once around the S^1 direction, and b is another loop near $\mathbb{T}^2$ that wraps once around the S^1-direction of $\mathbb{D}^2 \times S^1$, which is the other dimension of $\mathbb{T}^2 = S^1 \times S^1$. The interesting part is c, as it is represented by a loop in $\mathbb{T}^2$ that is parallel to $K_{p,q}$, thus it wraps p times around the direction of a and q times around the direction of b. This means $(j_A)_* c = a^p$ and $(j_B)_* c = b^q$, so putting all of this together yields:

Theorem 14.16. $\pi_1(S^3 \backslash K_{p,q}) \cong \{a, b \mid a^p = b^q\}$. ☐

Example 14.17. For $(p, q) = (1, 0)$, we obtain the knot group of the unknot: $\pi_1(S^3 \backslash K_{1,0}) \cong \{a, b \mid a = e\} = \{b \mid \varnothing\} = \mathbb{Z}$. In particular, this is an abelian group.

Example 14.18. The knot group of the trefoil is $\pi_1(S^3 \backslash K_{2,3}) \cong \{a, b \mid a^2 = b^3\}$. We proved in Exercise 12.20 that this group is not abelian, in contrast to Example 14.17, hence $\pi_1(S^3 \backslash K_{2,3})$ and $\pi_1(S^3 \backslash K_{1,0})$ are not isomorphic.

Corollary 14.19. *The knot complements* $\mathbb{R}^3 \backslash K_{1,0}$ *and* $\mathbb{R}^3 \backslash K_{2,3}$ *are not homeomorphic.* ☐

Before moving on[19] from the Seifert-van Kampen theorem, I would like to sketch one more application, which answers the question, "which groups can be fundamental groups of nice spaces?" If we are only interested in finitely-presented groups and decide that "nice" should mean "compact and Hausdorff", then the answer turns out to be that there is no restriction at all.

Theorem 14.20. *Every finitely-presented group is the fundamental group of some compact Hausdorff space.*

Proof. The following lemma will be used as an inductive step. Suppose X_0 is a compact Hausdorff space with a finitely-presented fundamental group

$$\pi_1(X_0, p) \cong \left\{\{a_i\} \,\middle|\, \{R_j\}\right\}.$$

Then for any loop $\gamma : (S^1, 1) \to (X_0, p)$, we claim that the space

$$X := \mathbb{D}^2 \cup_\gamma X_0 := \left(\mathbb{D}^2 \amalg X_0\right) \Big/ z \sim \gamma(z) \in X_0 \text{ for all } z \in \partial\mathbb{D}^2$$

is compact and Hausdorff with

$$\pi_1(X, p) \cong \left\{\{a_i\} \,\middle|\, \{R_j\},\ [\gamma] = e\right\},$$

i.e. its fundamental group has the same generators and one new relation, defined by setting $[\gamma] \in \pi_1(X_0, p)$ equal to the trivial element. This claim follows easily[20] from the Seifert-van Kampen

---

[19] We ran out of time in the actual lecture before we could talk about Theorem 14.20, but I am including it in the notes just because it is interesting.

[20] I am glossing over the detail where we need to prove that X is also compact and Hausdorff. This is not completely obvious, but it is yet another exercise in point-set topology that I feel justified in not explaining now that that portion of the course is finished.

theorem using the decomposition $X = A \cup B$ where $A = \mathring{\mathbb{D}}^2$ and B is an open neighborhood of X_0 obtained by adding a small annulus near the boundary of $\partial\mathbb{D}^2$. Since the annulus admits a deformation retraction to $\partial\mathbb{D}^2$, we have $B \underset{h.e.}{\simeq} X_0$, while $A \cap B \underset{h.e.}{\simeq} S^1$ and A is contractible. According to Corollary 12.19, $\pi_1(X, p)$ then inherits all the generators and relations of $\pi_1(B) \cong \pi_1(X_0)$, no new generators from $\pi_1(A) = 0$, and one new relation from the generator of $\pi_1(A \cap B) \cong \mathbb{Z}$, whose inclusion into A is trivial, so the relation says that its inclusion into B must become the trivial element. That inclusion is precisely $[\gamma] \in \pi_1(X_0, p)$, hence the claim is proved.

Now suppose G is a finitely-presented group with generators $x_1, \ldots, x_N$ and relations $w_1 = e, \ldots, w_m = e$ for $w_i \in F_{\{x_1,\ldots,x_N\}}$. We start with a space X_0 whose fundamental group is the free group on $\{x_1, \ldots, x_N\}$: the wedge sum of N circles will do. As the previous paragraph demonstrates, we can then attach a 2-disk for each individual relation we would like to add to the fundamental group, and doing this finitely many times produces a compact Hausdorff space with the desired fundamental group. □

15. Covering spaces and the lifting theorem

We now leave the Seifert-van Kampen theorem behind and introduce the second major tool for computing fundamental groups: the theory of covering spaces.

Definition 15.1. A map $f : Y \to X$ is called a **covering map** (*Überlagerung*), or simply a **cover** of X, if for every $x \in X$, there exists an open neighborhood $\mathcal{U} \subset X$ such that

$$f^{-1}(\mathcal{U}) = \bigcup_{\alpha \in J} \mathcal{V}_\alpha$$

for a collection of disjoint open subsets $\{\mathcal{V}_\alpha \subset Y\}_{\alpha \in J}$ such that $f|_{\mathcal{V}_\alpha} : \mathcal{V}_\alpha \to \mathcal{U}$ is a homeomorphism for each $\alpha \in J$. The domain Y of this map is called a **covering space** (*Überlagerungsraum*) of X. Any subset $\mathcal{U} \subset X$ satisfying the conditions stated above is said to be **evenly covered**.

Example 15.2. The map $f : \mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$ is a covering map of S^1.

Example 15.3. The map $S^1 \to S^1$ sending $e^{i\theta}$ to $e^{ki\theta}$ for any nonzero $k \in \mathbb{Z}$ is also a covering map of S^1.

Example 15.4. The **n-dimensional torus** $\mathbb{T}^n := \underbrace{S^1 \times \ldots \times S^1}_{n}$ admits a covering map

$$\mathbb{R}^n \to \mathbb{T}^n : (\theta_1, \ldots, \theta_n) \mapsto (e^{i\theta_1}, \ldots, e^{i\theta_n}).$$

More generally, it is straightforward to show that given any two covering maps $f_i : Y_i \to X_i$ for $i = 1, 2$, there is a "product" cover

$$Y_1 \times Y_2 \xrightarrow{f_1 \times f_2} X_1 \times X_2 : (x_1, x_2) \mapsto (f_1(x_1), f_2(x_2)).$$

Example 15.5. For any space X, the identity map $X \to X$ is trivially a covering map.

Example 15.6. Another trivial example of a covering map can be defined for any space X and any set J by setting $X_\alpha := X$ for every $\alpha \in J$ and defining $f : \coprod_{\alpha \in J} X_\alpha \to X$ as the unique map that restricts to each $X_\alpha = X$ as the identity map on X. This is a *disconnected* covering map. We will usually restrict our attention to covering spaces that are connected.

Example 15.7. For each $n \in \mathbb{N}$, the quotient projection $S^n \to \mathbb{RP}^n = S^n/\sim$ is a covering map.

Theorem 15.8. *If X is connected and $f : Y \to X$ is a cover, then the number (finite or infinite) of points in $f^{-1}(x) \subset Y$ does not depend on the choice of a point $x \in X$.*

Proof. Given $x \in X$, choose an evenly covered neighborhood $\mathcal{U} \subset X$ of x and write $f^{-1}(\mathcal{U}) = \bigcup_{\alpha \in J} \mathcal{V}_\alpha$. Then for every $y \in \mathcal{U}$, $|f^{-1}(y)| = |J|$, and it follows that for every $n \in \{0, 1, 2, 3, \ldots, \infty\}$, the subset $X_n := \{x \in X \mid |f^{-1}(x)| = n\} \subset X$ is open. If $x \in X_n$, notice that $\bigcup_{m \neq n} X_m$ is also open, thus X_n is also closed, so connectedness implies $X_n = X$. $\square$

In the setting of the above theorem, the number of points in $f^{-1}(x)$ is called the **degree** (*Grad*) of the cover. If $\deg(f) = n$, we sometimes call f an n-**fold** cover.

Examples 15.9. The cover $S^1 \to S^1 : z \mapsto z^k$ from Example 15.3 has degree $|k|$, while the quotient projection $S^n \to \mathbb{RP}^n$ has degree 2 and the cover $\mathbb{R} \to S^1$ from Example 15.2 has infinite degree.

Remark 15.10. Some authors strengthen the definition of a covering map $f : Y \to X$ by requiring f to be surjective. We did not require this in Definition 15.1, but notice that if X is connected, then it follows immediately from Theorem 15.8. In practice, it is only sensible to consider covers of connected spaces, and we shall always assume connectedness.

Note that in Definition 15.1, one should explicitly require the sets $\mathcal{V}_\alpha \subset f^{-1}(\mathcal{U})$ to be open. This is important, as part of the point of that definition is that X can be covered by open neighborhoods $\mathcal{U}$ whose preimages are homeomorphic to *disjoint unions* of copies of $\mathcal{U}$, i.e.

$$f^{-1}(\mathcal{U}) \cong \coprod_{\alpha \in J} \mathcal{U}.$$

This is true specifically because each of the sets $\mathcal{V}_\alpha$ is open, and therefore (as the complement of $\bigcup_{\beta \neq \alpha} \mathcal{V}_\beta$) also closed in $f^{-1}(\mathcal{U})$. To put it another way, in a covering map, every point $x \in X$ has a neighborhood $\mathcal{U}$ such that $f^{-1}(\mathcal{U})$ is the disjoint union of homeomorphic neighborhoods of the individual points in $f^{-1}(x)$. An important consequence of this definition is that every covering map $f : Y \to X$ is also a *local homeomorphism*, meaning that for each $y \in Y$ and $x := f(y)$, f maps some neighborhood of y homeomorphically to some neighborhood of x.

Almost every result in covering space theory is based on the answer to the following question: given a map $f : A \to X$ and a covering map $p : Y \to X$, can f be "lifted" to a map $\tilde{f} : A \to Y$ satisfying $p \circ \tilde{f} = f$? This problem can be summarized with the diagram

$$\begin{tikzcd} & Y \arrow[d, "p"] \\ A \arrow[r, "f"] \arrow[ur, dashed, "\tilde{f}"] & X \end{tikzcd} \tag{15.1}$$

in which the maps f and p are given, but the dashed arrow for $\tilde{f}$ indicates that we do not know whether such a map exists. If it does, then we call $\tilde{f}$ a **lift** of f to the cover. It is easy to see that lifts do not always exist: take for instance the cover $p : \mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$ and let $f : S^1 \to S^1$ be the identity map. A lift $\tilde{f} : S^1 \to \mathbb{R}$ would need to associate to every $e^{i\theta} \in S^1$ some point $\phi := \tilde{f}(e^{i\theta})$ such that $e^{i\phi} = e^{i\theta}$. It is easy to define a function that does this, but can we make it *continuous*? If it were continuous, then $\tilde{f}(e^{i\theta})$ would have to increase by 2π as $e^{i\theta}$ turns around the circle from $\theta = 0$ to $\theta = 2\pi$, producing *two* values $\tilde{f}(e^{2\pi i}) = \tilde{f}(1) + 2\pi$ even though $e^{2\pi i} = 1$. The goal for the remainder of this lecture is to determine precisely which maps can be lifted to which covering spaces and which cannot.

We start with the following observation: choose base points $a \in A$ and $x \in X$ to make $f : (A, a) \to (X, x)$ into a pointed map. Then if a lift $\tilde{f} : A \to Y$ exists and we set $y := \tilde{f}(a)$ to make $\tilde{f}$ a pointed map, p now becomes one as well since $p(y) = p(\tilde{f}(a)) = f(a) = x$, hence (15.1) becomes a diagram of pointed maps and induces a corresponding diagram of group homomorphisms

(15.2)
$$\begin{tikzcd} & \pi_1(Y,y) \arrow[d, "p_*"] \\ \pi_1(A,a) \arrow[ur, "\tilde{f}_*"] \arrow[r, "f_*"] & \pi_1(X,x). \end{tikzcd}$$

The existence of this diagram implies a nontrivial condition that relates the homomorphisms f_* and p_* but has nothing intrinsically to do with the lift: it implies $\operatorname{im} f_* \subset \operatorname{im} p_*$, i.e. these are two subgroups of $\pi_1(X,x)$, and one of them must be contained in the other. The lifting theorem states that under some assumptions that are satisfied by most reasonable spaces, this necessary condition is also sufficient.

THEOREM 15.11 (lifting theorem). *Assume X, Y, A are all path-connected spaces, A is also locally path-connected, $p : Y \to X$ is a covering map and $f : (A, a_0) \to (X, x_0)$ is a base-point preserving map. Then for any choice of base point $y_0 \in f^{-1}(x_0) \subset Y$, f admits a base-point preserving lift $\tilde{f} : (A, a_0) \to (Y, y_0)$ if and only if*

$$f_*\left(\pi_1(A,a_0)\right) \subset p_*\left(\pi_1(Y,y_0)\right),$$

and the point $y_0 = \tilde{f}(a_0)$ uniquely determines the lift $\tilde{f}$.

Let us discuss some applications before we get to the proof.

COROLLARY 15.12. *For any covering map $p : Y \to X$ between path-connected spaces and any space A that is simply connected and locally path-connected, every map $f : A \to X$ can be lifted to Y.* $\square$

COROLLARY 15.13. *For every base-point preserving covering map $p : (Y, y_0) \to (X, x_0)$ between path-connected spaces, the homomorphism $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ is injective.*

PROOF. Suppose $\tilde{\gamma} : (S^1, 1) \to (Y, y_0)$ is a loop such that $p_*[\tilde{\gamma}] = e \in \pi_1(X, x_0)$. Then $\gamma := p \circ \tilde{\gamma} : (S^1, 1) \to (X, x_0)$ admits an extension $u : (\mathbb{D}^2, 1) \to (X, x_0)$ with $u|_{\partial\mathbb{D}^2} = \gamma$. But $\mathbb{D}^2$ is simply connected, so u admits a lift $\tilde{u} : (\mathbb{D}^2, 1) \to (Y, y_0)$ satisfying $p \circ \tilde{u} = u$, thus $p \circ \tilde{u}|_{\partial\mathbb{D}^2} = \gamma$ implies that $\tilde{u}|_{\partial\mathbb{D}^2} : (S^1, 1) \to (Y, y_0)$ is a lift of γ. Uniqueness of lifts then implies $\tilde{u}|_{\partial\mathbb{D}^2} = \tilde{\gamma}$ and thus $[\tilde{\gamma}] = e \in \pi_1(Y, y_0)$. $\square$

COROLLARY 15.14. *If X is simply connected, then every path-connected covering space of X is also simply connected.* $\square$

EXAMPLE 15.15. Corollary 15.14 implies that there does not exist any covering map $S^1 \to \mathbb{R}$.

Here is an application important in complex analysis. Observe that

$$p : \mathbb{C} \to \mathbb{C}^* := \mathbb{C}\backslash\{0\} : z \mapsto e^z$$

is a covering map. Writing $p(x + iy) = e^x e^{iy}$, we can picture p as a transformation from Cartesian to polar coordinates: it maps every horizontal line $\{\operatorname{Im} z = \text{const}\}$ to a ray in $\mathbb{C}^*$ emanating from the origin, and every vertical line $\{\operatorname{Re} z = \text{const}\}$ to a circle in $\mathbb{C}^*$, which it covers infinitely many times. This shows that p is not bijective, so it has no global inverse, but it will admit inverses if we restrict it to suitably small domains, and it is useful to know what domains will generally suffice for this. In other words, we would like to know which open subsets $\mathcal{U} \subset \mathbb{C}^*$ can be the domain of a continuous function

$$\log : \mathcal{U} \to \mathbb{C} \quad \text{such that} \quad e^{\log z} = z \text{ for all } z \in \mathcal{U}.$$

For simplicity, we will restrict our attention to path-connected[21] domains and also assume $1 \in \mathcal{U}$, so that we can adopt the convention $\log(1) := 0$. Defining $f : (\mathcal{U}, 1) \hookrightarrow (\mathbb{C}^*, 1)$ as the inclusion, the desired function $\log : (\mathcal{U}, 1) \to (\mathbb{C}, 0)$ will then be the unique solution to the lifting problem

$$\begin{tikzcd} & (\mathbb{C}, 0) \arrow[d, "p"] \\ (\mathcal{U}, 1) \arrow[ur, dashed, "\log"] \arrow[r, hook, "f"] & (\mathbb{C}^*, 1) \end{tikzcd}$$

Theorem 15.11 now gives the answer: $\log : \mathcal{U} \to \mathbb{C}$ exists if and only if $f_*(\pi_1(\mathcal{U}, 1)) \subset p_*(\pi_1(\mathbb{C}, 0)) = 0$, or in other words, if every loop in $\mathcal{U}$ can be extended to a map $\mathbb{D}^2 \to \mathbb{C}^*$. Using the notion of the *winding number* from Exercise 10.27, this is the same as saying every loop $\gamma : S^1 \to \mathcal{U}$ satisfies $\operatorname{wind}(\gamma; 0) = 0$. For example, $\log : \mathcal{U} \to \mathbb{C}$ can be defined whenever $\mathcal{U}$ is simply connected, or if $\mathcal{U}$ has the shape of an annulus whose outer circle does not enclose the origin. Examples that do not work include any annulus whose inner circle encloses the origin: this will always contain a loop that winds nontrivially around the origin, so that trying to define log along this loop produces a function that shifts by $2\pi i$ as one rotates fully around the loop. Notice that when $\log : \mathcal{U} \to \mathbb{C}$ exists, it is uniquely determined by the condition $\log(1) = 0$; without this one could equally well modify any given definition of log by adding integer multiples of $2\pi i$.

The proof of the lifting theorem requires two lemmas that are also special cases of the theorem. We assume for the remainder of this lecture that $(Y, y_0) \xrightarrow{p} (X, x_0)$ is a covering map and X, Y and A are all path-connected.

LEMMA 15.16 (the path lifting property). *Every path $\gamma : (I, 0) \to (X, x_0)$ has a unique lift $\tilde{\gamma} : (I, 0) \to (Y, y_0)$.*

PROOF. Since I is compact, we can find a finite partition $0 =: t_0 < t_1 < \ldots < t_{N-1} < t_N := 1$ such that for each $j = 1, \ldots, N$, the image of $\gamma_j := \gamma|_{[t_{j-1}, t_j]}$ lies in an evenly covered open subset $\mathcal{U}_j \subset X$ with $p^{-1}(\mathcal{U}_j) = \bigcup_{\alpha \in J} \mathcal{V}_\alpha$. Now given any $y \in p^{-1}(\gamma(t_{j-1}))$, we have $y \in \mathcal{V}_\alpha$ for a unique $\alpha \in J$, and γ_j has a unique lift $\tilde{\gamma}_j : [t_{j-1}, t_j] \to Y$ with $\tilde{\gamma}_j(t_{j-1}) = y$, defined by

$$\tilde{\gamma}_j = (p|_{\mathcal{V}_\alpha})^{-1} \circ \gamma_j.$$

With this understood, the unique lift $\tilde{\gamma}$ of γ with $\tilde{\gamma}(0) = y_0$ can be constructed by lifting $\tilde{\gamma}_1$ as explained above, then lifting $\tilde{\gamma}_2$ with starting point $\tilde{\gamma}_2(t_1) := \tilde{\gamma}_1(t_1)$, and continuing in this way to cover the entire interval. □

LEMMA 15.17 (the homotopy lifting property). *Suppose $H : I \times A \to X$ is a homotopy with $H(0, \cdot) = f : A \to X$, and $\tilde{f} : A \to Y$ is a lift of f. Then there exists a unique lift $\widetilde{H} : I \times A \to Y$ of H satisfying $\widetilde{H}(0, \cdot) = \tilde{f}$.*

PROOF. The previous lemma implies that each of the paths $s \mapsto H(s, a) \in X$ for $a \in A$ have unique lifts $s \mapsto \widetilde{H}(s, a) \in Y$ with $\widetilde{H}(0, a) = \tilde{f}(a)$. One should then check that the map $\widetilde{H} : I \times A \to Y$ defined in this way is continuous; I leave this as an exercise. □

PROOF OF THEOREM 15.11. We shall first define an appropriate map $\tilde{f} : A \to Y$ and then show that the definition is independent of choices. Its uniqueness will be immediately clear, but its continuity will not be: in the final step we will use the hypothesis that A is locally path-connected in showing that $\tilde{f}$ is continuous.

---

[21] Since $\mathcal{U} \subset \mathbb{C}^*$ is open, it is locally path-connected, thus it will automatically be path-connected if it is connected.

Given $a \in A$, choose a path $a_0 \overset{\alpha}{\rightsquigarrow} a$, giving a path $x_0 \overset{f\circ\alpha}{\rightsquigarrow} f(a)$, which lifts via Lemma 15.16 to a unique path $\widetilde{f \circ \alpha}$ in Y that starts at y_0. If a lift $\tilde{f}$ exists, it clearly must satisfy

$$\tilde{f}(a) = \widetilde{f \circ \alpha}(1).$$

We claim that this point in Y does not depend on the choice of the path α, and thus gives a well-defined (though not necessarily continuous) map $\tilde{f} : A \to Y$. Indeed, suppose $a_0 \overset{\beta}{\rightsquigarrow} a$ is another path. Then $\alpha \cdot \beta^{-1}$ is a loop based at a_0 and thus represents an element of $\pi_1(A, a_0)$, and $f_*[\alpha \cdot \beta^{-1}] \in \pi_1(X, x_0)$ is represented by the loop $(f \circ \alpha) \cdot (f \circ \beta^{-1})$. The hypothesis $\operatorname{im} f_* \subset \operatorname{im} p_*$ then implies the existence of a loop $y_0 \overset{\tilde{\gamma}}{\rightsquigarrow} y_0$ in Y such that

$$[(f \circ \alpha) \cdot (f \circ \beta^{-1})] = p_*[\tilde{\gamma}] = [p \circ \tilde{\gamma}],$$

so there is a homotopy $H : I^2 \to X$ with $H(0, \cdot) = \gamma := p \circ \tilde{\gamma}$, $H(1, \cdot) = (f \circ \alpha) \cdot (f \circ \beta^{-1})$, and $H(s, 0) = H(s, 1) = x_0$ for all $s \in I$. Notice that $\tilde{\gamma}$ is a lift of $\gamma : (I, 0) \to (X, x_0)$. Now Lemma 15.17 provides a lift $\widetilde{H} : I^2 \to Y$ of H with $\widetilde{H}(0, \cdot) = \tilde{\gamma}$. In this homotopy, the paths $s \mapsto \widetilde{H}(s, 0)$ and $s \mapsto \widetilde{H}(s, 1)$ are lifts of the constant path $H(\cdot, 0) = H(\cdot, 1) \equiv x_0$ starting at $\tilde{\gamma}(0) = \tilde{\gamma}(1) = y_0$, so the uniqueness in Lemma 15.16 implies that both are also constant paths, hence $\widetilde{H}(s, 0) = \widetilde{H}(s, 1) = y_0$ for all $s \in I$. This shows that the unique lift of $(f \circ \alpha) \cdot (f \circ \beta^{-1})$ to a path in Y starting at y_0 is actually a loop, i.e. its end point is also y_0: indeed, this lift is $\widetilde{H}(1, \cdot)$. This lift is necessarily the concatenation of the lift $\widetilde{f \circ \alpha}$ of $f \circ \alpha$ starting at y_0 with the lift of $f \circ \beta^{-1}$ starting at $\widetilde{f \circ \alpha}(1)$. Since it ends at y_0, we conclude that this second lift is simply the inverse of $\widetilde{f \circ \beta}$, implying that

$$\widetilde{f \circ \alpha}(1) = \widetilde{f \circ \beta}(1),$$

which proves the claim.

It remains to show that $\tilde{f} : A \to Y$ as defined by the above procedure is continuous. Given $a \in A$ with $x = f(a) \in X$ and $y = \tilde{f}(a) \in Y$, choose any neighborhood $\mathcal{V} \subset Y$ of y that is small enough for $\mathcal{U} := p(\mathcal{V}) \subset X$ to be an evenly covered neighborhood of x, with $p|_{\mathcal{V}} : \mathcal{V} \to \mathcal{U}$ a homeomorphism. It will suffice to show that a has a neighborhood $\mathcal{O} \subset A$ with $\tilde{f}(\mathcal{O}) \subset \mathcal{V}$. Since A is locally path-connected, we can choose $\mathcal{O} \subset f^{-1}(\mathcal{U})$ to be a path-connected neighborhood of a, fix a path $a_0 \overset{\gamma}{\rightsquigarrow} a$ in A and, for any $a' \in \mathcal{O}$, choose a path $a \overset{\beta}{\rightsquigarrow} a'$ in $\mathcal{O}$. Now $\gamma \cdot \beta$ is a path from a_0 to a', so

$$\tilde{f}(a) = \widetilde{f \circ \gamma}(1) = y \in \mathcal{V} \quad \text{and} \quad \tilde{f}(a') = \widetilde{f \circ \gamma} \cdot \widetilde{f \circ \beta}(1),$$

where $\widetilde{f \circ \beta}$ is the unique lift of $f \circ \beta$ starting at y. Since $f \circ \beta$ lies entirely in the evenly covered neighborhood $\mathcal{U}$, this second lift is simply $(p|_{\mathcal{V}})^{-1} \circ (f \circ \beta)$, which lies entirely in $\mathcal{V}$, proving $\tilde{f}(a') \in \mathcal{V}$. □

EXAMPLE 15.18. If the local path-connectedness assumption on A is dropped, then the proof above gives a procedure for defining a unique lift $\tilde{f} : A \to Y$, but it may fail to be continuous. A concrete example is depicted in [**Hat02**, p. 79], Exercise 7. The idea is to define A as a space that mostly consists of the usual circle $S^1 \subset \mathbb{R}^2$, but replace a portion just to the right of the top point $(0, 1)$ with a curve resembling the graph of the function $y = \sin(1/x) + 1$. The point $(0, 1)$ is included in A, along with every point of the usual circle just to the left of it, but on the right, A consists of an infinitely long curve that is compressed into a compact space and has accumulation points along an interval but no well-defined limit. This space is path-connected, because one can start from $(0, 1)$ and go around the circle to reach any other point, including any point on the infinitely long compressed sine curve; it is also simply connected, due to the fact that continuous paths along the compressed sine curve can never actually reach the end of it, but must instead go back the other way around the circle before they can reach $(0, 1)$. But A is not locally path-connected, because sufficiently small neighborhoods of $(0,1)$ in A always contain many disjoint segments of the compressed sine curve and thus cannot be path-connected. Now consider the covering map $\mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$ and a continuous map $f : A \to S^1$ defined as the identity on most of A, but projecting the graph of $y = \sin(1/x) + 1$ to the circle in the obvious way near $(0,1)$. One can define a lift $\tilde{f} : A \to \mathbb{R}$ by choosing $\tilde{f}(0,1)$ to be any point in $p^{-1}(f(0,1))$ and then lifting paths to define $\tilde{f}$ everywhere else. But since every neighborhood of $(0,1)$ contains some points that cannot be reached except by paths rotating almost all the way around the circle, this neighborhood will contain points $a \in A$ for which $\tilde{f}(a)$ differs from $\tilde{f}(0,1)$ by nearly 2π. In particular, $\tilde{f}$ cannot be continuous at $(0,1)$.

16. Classification of covers

Throughout this lecture, all spaces should be assumed path-connected and locally path-connected unless otherwise noted. We will occasionally need a slightly stronger condition, which we will abbreviate with the word "reasonable":[22]

Definition 16.1. We will say that a space X is **reasonable** if it is path-connected and locally path-connected, and every point $x \in X$ has a simply connected neighborhood.

For the purposes of the theorems in this lecture, the definition of the term "reasonable" can be weakened somewhat at the expense of making it more complicated, but we will stick with the above definition since it is satisfied by almost all spaces we would ever like to consider. A popular example of an "unreasonable" space is the so-called *Hawaiian earring*, see Exercise 13.2(c).

We will state several theorems in this lecture related to the problem of classifying covers of a given space. All of them are in some way applications of the lifting theorem (Theorem 15.11). Before stating them, we need to establish what it means for two covers of the same space to be "equivalent".

Definition 16.2. Given two covers $p_i : Y_i \to X$ for $i = 1, 2$, a **map of covers** from p_1 to p_2 is a map $f : Y_1 \to Y_2$ such that $p_2 \circ f = p_1$, i.e. the following diagram commutes:

$$\begin{tikzcd} Y_1 \arrow[rr, "f"] \arrow[dr, "p_1"'] & & Y_2 \arrow[dl, "p_2"] \\ & X & \end{tikzcd} \tag{16.1}$$

Additionally, we call f an **isomorphism of covers** if there also exists a map of covers from p_2 to p_1 that inverts f; this is true if and only if the map $f : Y_1 \to Y_2$ is a homeomorphism, since its inverse $f^{-1} : Y_2 \to Y_1$ is then automatically a map of covers from p_2 to p_1. If such an isomorphism exists, we say that the two covers p_1 and p_2 are **isomorphic** (or **equivalent**). If base points $x \in X$ and $y_i \in Y_i$ are specified such that $p_i : (Y_i, y_i) \to (X, x)$ and $f : (Y_1, y_1) \to (Y_2, y_2)$ are also pointed maps, then we call f an **isomorphism of pointed covers**. In the case where p_1 and p_2 are both the same cover $p : Y \to X$, an isomorphism of covers from p to itself is called a **deck transformation**[23] (*Decktransformation*) of $p : Y \to X$.

The terms **covering translation** and **automorphism** are also sometimes used as synonyms for "deck transformation". The set of all deck transformations of a given cover $p : Y \to X$ forms a

---

[22] This is not a universally standard term.

[23] This terminology gives you a hint that some portion of this subject was developed by German mathematicians in the time before English was fully established as an international language. I don't happen to know who invented the term.

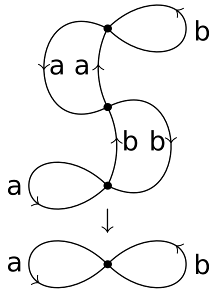


FIGURE 10. A 3-fold cover of $S^1 \vee S^1$ with trivial automorphism group.

group, called the **automorphism group**

$$\mathrm{Aut}(p) := \left\{ f : Y \to Y \;\middle|\; f \text{ is a homeomorphism such that } p \circ f = p \right\},$$

where the group operation is defined by composition of maps.

EXAMPLE 16.3. For the cover $p : \mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$, $\mathrm{Aut}(p)$ consists of all maps $f_k : \mathbb{R} \to \mathbb{R}$ of the form $f_k(\theta) = \theta + 2\pi k$ for $k \in \mathbb{Z}$, so in particular, $\mathrm{Aut}(p)$ is isomorphic to $\mathbb{Z}$.

EXAMPLE 16.4. Figure 10 illustrates a covering map $p : Y \to S^1 \vee S^1$ of degree 3. If we label the base point of $S^1 \vee S^1$ as x, then the three elements of $p^{-1}(x) \subset Y$ are the three dots in the top portion of the diagram: label them y_1, y_2 and y_3 from bottom to top. The covering map is defined such that each loop or path beginning and ending at any of the points y_1, y_2, y_3 is sent to the loop in $S^1 \vee S^1$ labeled by the same letter with the orientations of the arrows matching. Suppose $f : Y \to Y$ is a deck transformation satisfying $f(y_1) = y_2$. Then since f is a homeomorphism, it must map the loop labeled a based at y_1 to a loop based at y_2 that also must be labeled a. But no such loop exists, so we conclude that there is no deck transformation sending y_1 to y_2. By similar arguments, it is not hard to show that the only deck transformation of this cover is the identity map, in other words, $\mathrm{Aut}(p)$ is the trivial group.

Almost everything we will be able to prove about maps of covers is based on the following observation: if the diagram (16.1) commutes, it means that $f : Y_1 \to Y_2$ is a *lift* of the map $p_1 : Y_1 \to X$ to the cover Y_2, i.e. in our previous notation for lifts, $f = \tilde{p}_1$. The fact that p_1 itself is a covering map is irrelevant for this observation. Now if all the spaces involved are path-connected and locally path-connected, the lifting theorem gives us a condition characterizing the existence and uniqueness of a map of covers: for any choices of base points $x \in X$, $y_1 \in p_1^{-1}(x) \subset Y_1$ and $y_2 \in p_2^{-1}(x) \subset Y_2$, a map of covers $f : Y_1 \to Y_2$ satisfying $f(y_1) = y_2$ exists (and is unique) if and only if

$$(p_1)_*\pi_1(Y_1, y_1) \subset (p_2)_*\pi_1(Y_2, y_2).$$

This map will then be an isomorphism if and only if there exists a map of covers going the other direction, and the latter exists if and only if the reverse inclusion holds. This proves:

THEOREM 16.5. *Two covers $p_i : Y_i \to X$ for $i = 1, 2$ are isomorphic if and only if for some choice of base points $x \in X$ and $y_i \in p_i^{-1}(x) \subset Y_i$ for $i = 1, 2$, the subgroups $(p_1)_*\pi_1(Y_1, y_1)$ and $(p_2)_*\pi_1(Y_2, y_2)$ in $\pi_1(X, x)$ are identical.* □

Next we use the same perspective to study deck transformations of a single cover $p : Y \to X$. Given $x \in X$ and $y_1, y_2 \in p^{-1}(x) \subset Y$, the uniqueness of lifts implies that there exists at most one deck transformation $f : Y \to Y$ sending y_1 to y_2. We've seen in Example 16.4 that this transformation might not always exist.

Definition 16.6. A cover $p : Y \to X$ is called **regular** (or equivalently **normal**) if for every $x \in X$ and all $y_1, y_2 \in p^{-1}(x) \subset Y$, there exists a deck transformation sending y_1 to y_2.

The following exercise says that in order to check whether a cover of a path-connected space is regular, it suffices to choose a base point $x \in X$ and investigate whether deck transformations can be used to relate arbitrary points in the preimage of *that particular point.* The proof is an easy application of the path lifting property (Lemma 15.16).

Exercise 16.7. Show that if $p : Y \to X$ is a covering map and X is path-connected, then p is also regular if the following slightly weaker condition holds: for some fixed $x \in X$, any two elements $y_1, y_2 \in p^{-1}(x) \subset X$ satisfy $y_2 = f(y_1)$ for some deck transformation $f \in \mathrm{Aut}(p)$.

If $\deg(p) < \infty$, the previous remarks about uniqueness of deck transformations imply $|\mathrm{Aut}(p)| \leqslant \deg(p)$, and equality is satisfied if and only if p is regular. By the lifting theorem, the desired deck transformation sending y_1 to y_2 will exist if and only if

$$p_*\pi_1(Y, y_1) = p_*\pi_1(Y, y_2). \tag{16.2}$$

Let us try to translate this into a condition for recognizing when p is regular. Recall that any path $y_1 \overset{\tilde{\gamma}}{\rightsquigarrow} y_2$ in Y determines an isomorphism

$$\Phi_{\tilde{\gamma}} : \pi_1(Y, y_2) \to \pi_1(Y, y_1) : [\alpha] \mapsto [\tilde{\gamma} \cdot \alpha \cdot \tilde{\gamma}^{-1}].$$

Since y_1 and y_2 are both in $p^{-1}(x)$, the projection of this concatenation down to X gives a concatenation of *loops*, i.e. $\gamma := p \circ \tilde{\gamma}$ is a loop $x \rightsquigarrow x$ and thus represents an element $[\gamma] \in \pi_1(X, x)$. Now in order to check whether (16.2) holds, we can represent an arbitrary element of $\pi_1(Y, y_1)$ as $\Phi_{\tilde{\gamma}}[\alpha]$ for some loop $y_2 \overset{\alpha}{\rightsquigarrow} y_2$, and then observe

$$p_*\Phi_{\tilde{\gamma}}[\alpha] = [p \circ (\tilde{\gamma} \cdot \alpha \cdot \tilde{\gamma}^{-1})] = [\gamma \cdot (p \circ \alpha) \cdot \gamma^{-1}] = [\gamma] p_*[\alpha][\gamma]^{-1}.$$

This proves that the subgroup $p_*\pi_1(Y, y_1) \subset \pi_1(X, x)$ is the conjugate of $p_*\pi_1(Y, y_2) \subset \pi_1(X, x)$ by the specific element $[\gamma] \in \pi_1(X, x)$, so the desired deck transformation exists if and only if $p_*\pi_1(Y, y_2)$ is invariant under conjugation with $[\gamma]$. We could now ask the same question about deck transformations sending y_i to y_2 for arbitrary $y_i \in p^{-1}(x)$, and the answer in each case can be expressed in terms of conjugation of $p_*\pi_1(Y, y_2)$ by some element $[\gamma] \in \pi_1(X, x)$ for which the loop γ lifts to a path $y_i \overset{\tilde{\gamma}}{\rightsquigarrow} y_2$. Now observe: *any* loop $x \overset{\gamma}{\rightsquigarrow} x$ can arise in this way for some choice of $y_i \in p^{-1}(x)$. Indeed, if γ is given, then γ^{-1} has a unique lift to a path from y_2 to some other point in $p^{-1}(x)$, and the inverse of this path is then a lift of γ. Using Exercise 16.7 above, the question of regularity therefore reduces to the question of whether $p_*\pi_1(Y, y_2)$ is invariant under arbitrary conjugations, and we have thus proved:

Theorem 16.8. *If Y and X are path-connected and locally path-connected, then a cover $p : (Y, y_0) \to (X, x_0)$ is regular if and only if the subgroup $p_*\pi_1(Y, y_0) \subset \pi_1(X, x_0)$ is normal.* □

Notice that while the algebraic condition in this theorem appears to depend on a choice of base points, the condition of p being regular clearly does not. It follows that if $p_*\pi_1(Y, y_0) \subset \pi_1(X, x_0)$ is a normal subgroup, then this condition will remain true for any other choice of base points $x \in X$ and $y \in p^{-1}(x) \subset Y$.

The next two results require the restriction to "reasonable" spaces in the sense of Definition 16.1.

Theorem 16.9 (the Galois correspondence). *If X is a reasonable space with base point $x_0 \in X$, there is a natural bijection from the set of all isomorphism classes of pointed covers $p : (Y, y_0) \to (X, x_0)$ to the set of all subgroups of $\pi_1(X, x_0)$: it is defined by*

$$[p : (Y, y_0) \to (X, x_0)] \mapsto p_*\pi_1(Y, y_0).$$

It is easy to verify from the definition of isomorphism for covers that the map in this theorem is well defined, and we proved in Theorem 16.5 that it is injective. Surjectivity will be a consequence of the following result, which will be proved in the next lecture.

THEOREM 16.10. *Every reasonable space admits a simply connected covering space.*

Notice that if $p_i : (Y_i, y_i) \to (X, x_0)$ for $i = 1, 2$ are two reasonable covers satisfying $\pi_1(Y_1) = \pi_1(Y_2) = 0$, then Theorem 16.5 implies that they are isomorphic covers. For this reason it is conventional to abuse terminology slightly by referring to any simply connected cover of a given space X as "the" **universal cover** (*universelle Überlagerung*) of X. It is often denoted by $\widetilde{X}$.

EXAMPLES 16.11. The universal cover $\widetilde{S^1}$ of S^1 is $\mathbb{R}$, due to the covering map $\mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$. Similarly, $\widetilde{\mathbb{RP}^n} \cong S^n$ for $n \geqslant 2$, and $\widetilde{\mathbb{T}^n} \cong \mathbb{R}^n$.

A substantially less obvious class of examples is given by the surfaces Σ_g of genus $g \geqslant 2$: these have universal cover $\widetilde{\Sigma}_g \cong \mathbb{R}^2$. It would take us too far afield to explain why, but one standard way of constructing this cover comes from hyperbolic geometry, where instead of $\mathbb{R}^2$ we consider the open disk $\mathring{\mathbb{D}}^2$ with a Riemannian metric that has constant negative curvature. One can identify each of the surfaces Σ_g with the quotient of $\mathring{\mathbb{D}}^2$ by a suitable group of isometries and then define a covering map $\mathring{\mathbb{D}}^2 \to \Sigma_g$ as the quotient projection.

For the remainder of this lecture, fix a base-point preserving covering map $p : (Y, y_0) \to (X, x_0)$ where X and Y are assumed reasonable, and denote

$$G := \pi_1(X, x_0), \qquad H := p_*\pi_1(Y, y_0) \subset G.$$

If H is not a normal subgroup, then there is no natural notion of a quotient group G/H, but we can still define G/H as the *set* of left cosets

$$G\big/H = \{gH \subset G \mid g \in G\},$$

where gH denotes the subset $\{gh \mid h \in H\} \subset G$. One can similarly consider the set of right cosets

$$H\big\backslash G = \{Hg \subset G \mid g \in G\}.$$

These two sets are identical if and only if H is normal, in which case both are denoted by G/H and they form a group. With or without this condition, $G\big/H$ and $H\big\backslash G$ have the same number (finite or infinite) of elements, which is called the **index** of H in G and denoted by

$$[G : H] := \left|G\big/H\right| = \left|H\big\backslash G\right|.$$

In the following we will make repeated use of the fact that for any $y \in p^{-1}(x_0)$, any path $y_0 \overset{\tilde{\gamma}}{\rightsquigarrow} y$ gives rise to a loop $\gamma := p \circ \tilde{\gamma}$ based at x_0, and conversely, any such loop gives rise to a path that starts at y_0 and ends at some point in $p^{-1}(x_0)$.

LEMMA 16.12. *There is a natural bijection*

$$\Phi : p^{-1}(x_0) \to H\big\backslash G : y \mapsto H[\gamma],$$

where $x_0 \overset{\gamma}{\rightsquigarrow} x_0$ *is any loop that lifts to a path* $y_0 \overset{\tilde{\gamma}}{\rightsquigarrow} y$.

COROLLARY 16.13. $\deg(p) = [G : H]$. □

Proof of Lemma 16.12. We first show that Φ is well defined. Given two choices of paths $\tilde{\alpha}, \tilde{\beta}$ from y_0 to y, we have loops $\alpha := p \circ \tilde{\alpha}$ and $\beta := p \circ \tilde{\beta}$ based at x_0, and $\tilde{\alpha} \cdot \tilde{\beta}^{-1}$ is a loop based at y_0. We therefore have

$$[\alpha][\beta]^{-1} = [p \circ (\tilde{\alpha} \cdot \tilde{\beta}^{-1})] = p_*[\tilde{\alpha} \cdot \tilde{\beta}^{-1}] \in H,$$

implying $H[\alpha] = H[\beta]$.

The surjectivity of Φ is obvious: given $[\gamma] \in G$, there exists a lift $\tilde{\gamma}$ of γ to a path from y_0 to some point $y \in p^{-1}(x_0)$, so $\Phi(y) = H[\gamma]$.

To see that Φ is injective, suppose $\Phi(y) = \Phi(y')$, choose paths $y_0 \overset{\tilde{\alpha}}{\rightsquigarrow} y$ and $y_0 \overset{\tilde{\beta}}{\rightsquigarrow} y'$, giving rise to loops $\alpha := p \circ \tilde{\alpha}$ and $\beta := p \circ \tilde{\beta}$ based at x_0 such that

$$H[\alpha] = \Phi(y) = \Phi(y') = H[\beta],$$

thus $[\alpha][\beta]^{-1} \in H$. It follows that there exists a loop $y_0 \overset{\tilde{\gamma}}{\rightsquigarrow} y_0$ projecting to $\gamma := p \circ \tilde{\gamma}$ such that $[\alpha \cdot \beta^{-1}] = [\gamma]$, hence $[\alpha] = [\gamma] \cdot [\beta]$, so α is homotopic to $\gamma \cdot \beta$ with fixed end points. Since γ lifts to a loop $\tilde{\gamma}$ and homotopies can also be lifted, we conclude that $\tilde{\alpha}$ is homotopic to $\tilde{\gamma} \cdot \tilde{\beta}$ with fixed end points, implying $y = \tilde{\alpha}(1) = \tilde{\beta}(1) = y'$. □

If the cover is regular so $H \subset G$ is normal, then $\deg(p) = |\operatorname{Aut}(p)|$, and Corollary 16.13 therefore implies that $\operatorname{Aut}(p)$ has the same order as the quotient group G/H. The next result should then seem relatively unsurprising.

Theorem 16.14. *For a regular cover* $p : (Y, y_0) \to (X, x_0)$ *of reasonable spaces with* $\pi_1(X, x_0) = G$ *and* $p_*\pi_1(Y, y_0) = H \subset G$*, there exists a group isomorphism*

$$\Psi : \operatorname{Aut}(p) \to G/H : f \mapsto [\gamma]H,$$

where $x_0 \overset{\gamma}{\rightsquigarrow} x_0$ *is any loop that has a lift to a path from* y_0 *to* $f(y_0)$*.*

Notice that the universal cover $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ is automatically regular since the trivial subgroup of $\pi_1(X, x_0)$ is always normal, so applying this theorem to the universal cover gives:

Corollary 16.15. *For the universal cover* $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$*, there is an isomorphism* $\operatorname{Aut}(p) \to \pi_1(X, x_0)$ *sending each deck transformation* f *to the homotopy class of any loop* $x_0 \rightsquigarrow x_0$ *that lifts to a path* $\tilde{x}_0 \rightsquigarrow f(\tilde{x}_0)$*.* □

Proof of Theorem 16.14. Regularity implies that the map $\operatorname{Aut}(p) \to p^{-1}(x_0) : f \mapsto f(y_0)$ is bijective, so Ψ is then well defined and bijective due to Lemma 16.12. For the identity element $\operatorname{Id} \in \operatorname{Aut}(p)$, we have $\Psi(\operatorname{Id}) = [\gamma]H$ for any loop γ that lifts to a loop from y_0 to $\operatorname{Id}(y_0) = y_0$, which means $[\gamma] \in H$, so $[\gamma]H$ is the identity element in G/H.

It remains to show that $\Psi(f \circ g) = \Psi(f)\Psi(g)$ for any two deck transformations $f, g \in \operatorname{Aut}(p)$. Choose loops α, β based at x_0 which lift to paths $y_0 \overset{\tilde{\alpha}}{\rightsquigarrow} f(y_0)$ and $y_0 \overset{\tilde{\beta}}{\rightsquigarrow} g(y_0)$. Then $f \circ \tilde{\beta}$ is a path from $f(y_0)$ to $f \circ g(y_0)$ and can thus be concatenated with $\tilde{\alpha}$, forming a path

$$y_0 \overset{\tilde{\alpha} \cdot (f \circ \tilde{\beta})}{\rightsquigarrow} f \circ g(y_0).$$

Now since $f \in \operatorname{Aut}(p)$, $p \circ f = p$ implies $p \circ (f \circ \tilde{\beta}) = p \circ \tilde{\beta} = \beta$, thus

$$\Psi(f \circ g) = [p \circ (\tilde{\alpha} \cdot (f \circ \tilde{\beta}))] = [\alpha][\beta] = \Psi(f)\Psi(g).$$

□

Corollary 16.15 says that we can compute the fundamental group of any reasonable space X if we can understand the deck transformations of its universal cover. Combining this with the natural bijection $\mathrm{Aut}(p) \to p^{-1}(x_0)$ that sends each deck transformation to its image on the base point, we also obtain from this an intuitively appealing interpretation of the meaning of $\pi_1(X, x_0)$: every loop γ based at x_0 lifts uniquely to a path starting at $\tilde{x}_0$ and ending at some point in $p^{-1}(x_0)$. As far as $\pi_1(X, x_0)$ is concerned, all that matters is the end point of the lift: two loops are equivalent in $\pi_1(X, x_0)$ if and only if their lifts to $\widetilde{X}$ have the same end point, and a loop is trivial in $\pi_1(X, x_0)$ if and only if its lift to $\widetilde{X}$ is also a loop.

EXAMPLE 16.16. Applying Corollary 16.15 to the cover $p : \mathbb{R} \to S^1 : \theta \mapsto e^{i\theta}$ reproduces the isomorphism $\pi_1(S^1, 1) \cong \mathbb{Z}$ we discussed at the end of Lecture 9. The loop $\gamma_k(t) := e^{2\pi i k t}$ in S^1 for each $k \in \mathbb{Z}$ lifts to $\mathbb{R}$ with base point 0 as the path $\tilde{\gamma}_k(t) = 2\pi k t$.

EXAMPLE 16.17. For each $n \geqslant 2$, Corollary 16.15 implies $\pi_1(\mathbb{RP}^n) \cong \mathbb{Z}_2$, as this is the automorphism group of the universal cover $p : S^n \to \mathbb{RP}^n$, defined as the natural quotient projection. Concretely, after fixing base points $x_0 \in \mathbb{RP}^n$ and $y_0 \in p^{-1}(x_0) \subset S^n$, each loop in $\mathbb{RP}^n$ based at x_0 lifts to S^n as a path that starts at y_0 and ends at either y_0 or its antipodal point $-y_0$. The nontrivial element of $\pi_1(\mathbb{RP}^n, x_0)$ is thus represented by any loop whose lift to S^n starts and ends at antipodal points.

17. The universal cover and group actions

In Theorem 16.14, we saw a formula that can be used to compute the automorphism group of any regular cover as a quotient of two fundamental groups. I want to mention how this generalizes for non-regular covers, though I will leave most of the details as an exercise. One way to approach the problem is as follows: any pointed covering map $p : (Y, y_0) \to (X, x_0)$ of reasonable spaces can be fit into a diagram

$$\begin{tikzcd}(Z, z_0) \arrow[r, "q"] \arrow[rr, bend right, "P"'] & (Y, y_0) \arrow[r, "p"] & (X, x_0),\end{tikzcd} \tag{17.1}$$

in which q and P are also pointed covering maps and are both *regular*. For example, if you already believe that every reasonable space has a universal cover (and we shall prove this below), then we can always take $q : Z \to Y$ to be the universal cover of Y, which makes $P : Z \to X$ the universal cover of X since $\pi_1(Z) = 0$, and universal covers are always regular because the trivial subgroup is always normal. In this case, Corollary 16.15 gives us natural isomorphisms $\mathrm{Aut}(P) \cong \pi_1(X, x_0)$ and $\mathrm{Aut}(q) \cong \pi_1(Y, y_0)$. This is not true if Z is not simply connected, and we will not assume this in the following exercise, but it turns out that if P and q are nonetheless regular, then we can derive a formula for $\mathrm{Aut}(p)$ in terms of the other two automorphism groups.

EXERCISE 17.1. Assuming the spaces in (17.1) are all reasonable, let us abbreviate the automorphism groups of P and q by

$$G := \mathrm{Aut}(P), \qquad \text{and} \qquad H := \mathrm{Aut}(q).$$

(a) Use the path-lifting property to prove the following lemma: If $\Psi \in G$ and $\psi \in \mathrm{Aut}(p)$ are deck transformations for which the relation $q \circ \Psi = \psi \circ q$ holds at the base point $z_0 \in Z$, then it holds everywhere.
Hint: For any $z \in Z$, choose a path from z_0 to z, then use Ψ, ψ and the covering projections to cook up other paths in Z, Y and X. Some of them are lifts of others, and two important ones will turn out to be the same.

(b) Deduce from part (a) that H is the subgroup of G consisting of all deck transformations $\Psi : Z \to Z$ for P that satisfy $\Psi(z_0) \in q^{-1}(y_0)$.

(c) Show that if $P : Z \to X$ is regular then so is $q : Z \to Y$. Give two proofs: one using the result of part (b), and another using the characterization of regularity in terms of normal subgroups.

(d) The **normalizer** (*Normalisator*) $N(H) \subset G$ of the subgroup H is by definition the largest subgroup of G that contains H as a normal subgroup, i.e.

$$N(H) := \left\{ g \in G \;\middle|\; gHg^{-1} = H \right\}.$$

Show that if the cover $q : Z \to Y$ is regular, then for any $\Psi \in N(H)$, there exists a deck transformation $\psi : Y \to Y$ of p satisfying the relation $q \circ \Psi = \psi \circ q$, and it is unique. Moreover, the correspondence $\Psi \mapsto \psi$ defines a group homomorphism $N(H) \to \mathrm{Aut}(p)$ whose kernel is H.

(e) Show that if the cover $P : Z \to X$ is also regular, then the homomorphism $N(H) \to \mathrm{Aut}(p)$ in part (d) is also surjective, and thus descends to an isomorphism

$$N(H)/H \xrightarrow{\cong} \mathrm{Aut}(p).$$

Applying Exercise 17.1 with Z simply connected now gives:

COROLLARY 17.2. *For any covering map* $p : (Y, y_0) \to (X, x_0)$ *of reasonable spaces with* $\pi_1(X, x_0) = G$ *and* $p_* \pi_1(Y, y_0) = H \subset G$*, there is a natural isomorphism* $\mathrm{Aut}(p) \cong N(H)/H$. ☐

Notice that there always exists a subgroup of G in which H is normal, e.g. H itself is such a subgroup, and it may well happen that no larger subgroup satisfies this condition, in which case $N(H) = H$ and $\mathrm{Aut}(p)$ is therefore trivial. If H is normal in G, then $N(H) = G$ and the cover is therefore regular, hence Corollary 17.2 reduces to Theorem 16.14.

Moving on from non-regular covers, we have some unfinished business from the previous lecture: it remains to prove the surjectivity of the Galois correspondence (Theorem 16.9), and the existence of the universal cover (Theorem 16.10). The latter is actually a special case of the former: recall from Corollary 15.13 that the homomorphism $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ induced by a covering map $p : (Y, y_0) \to (X, x_0)$ is always injective, thus the existence of a universal cover amounts to the statement that the image of the Galois correspondence includes the trivial subgroup of $\pi_1(X, x_0)$. We will prove this first, and then use it to deduce the Galois correspondence in full generality.

As before, we need to restrict our attention to "reasonable spaces," meaning spaces that are path-connected and locally path-connected, and in which every point has a simply connected neighborhood. The first two conditions are needed in order to apply the lifting theorem, which we used several times in the previous lecture. The third condition has not yet been used, but this is the moment where we will need it. In constructing a universal cover $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$, the theorems at the end of the previous lecture give some useful intuition on what to aim for: in particular, there needs to be a one-to-one correspondence between $p^{-1}(x_0) \subset \widetilde{X}$ and $\pi_1(X, x_0)$. What we will actually construct is a cover for which these two sets are not just in bijective correspondence but are literally the same set. In set-theoretic terms, the construction is quite straightforward, but giving it a topology that makes it a covering map is a bit subtle—that is where we will need to assume that simply connected neighborhoods exist.

PROOF OF THEOREM 16.10 (THE UNIVERSAL COVER). We will not give every detail but sketch the main idea. Given a reasonable space X with base point $x_0 \in X$, define the set

$$\widetilde{X} := \{\text{paths } \gamma : (I, 0) \to (X, x_0)\} \Big/ \underset{h+}{\sim},$$

i.e. it is the set of all equivalence classes of paths that start at the base point, with equivalence defined as homotopy with fixed end points. Since this definition does not specify the end point of any path but the equivalence relation leaves these end points unchanged, we obtain a natural map

$$p : \widetilde{X} \to X : [\gamma] \mapsto \gamma(1),$$

which is obviously surjective since X is path-connected. Notice that $p^{-1}(x_0) = \pi_1(X, x_0)$.

We claim that $\widetilde{X}$ can be assigned a topology that makes $p : \widetilde{X} \to X$ into a covering map. To see this, suppose $\mathcal{U} \subset X$ is a path-connected subset and $i^{\mathcal{U}} : \mathcal{U} \hookrightarrow X$ denotes its inclusion. For any point $x \in \mathcal{U}$, the induced homomorphism $i^{\mathcal{U}}_* : \pi_1(\mathcal{U}, x) \to \pi_1(X, x)$ is trivial if and only if every loop $S^1 \to \mathcal{U}$ based at x can be extended to a map $\mathbb{D}^2 \to X$. Notice that this is weaker in general than demanding an extension $\mathbb{D}^2 \to \mathcal{U}$; the latter would mean that $\mathcal{U}$ is simply connected, but we do not want to assume this. Notice also that if this condition holds for some choice of base point $x \in \mathcal{U}$, then the usual change of base-point arguments imply that it will hold for any other base point $y \in \mathcal{U}$, thus we can sensibly speak of the condition that $i^{\mathcal{U}}_* : \pi_1(\mathcal{U}) \to \pi_1(X)$ is trivial. With this understood, consider the collection of sets

$$\mathcal{B} := \left\{ \mathcal{U} \subset X \;\middle|\; \mathcal{U} \text{ is open and path-connected and } i^{\mathcal{U}}_* : \pi_1(\mathcal{U}) \to \pi_1(X) \text{ is trivial} \right\}.$$

It is a straightforward exercise to verify the following properties:

1. $\mathcal{U} \in \mathcal{B}$ if and only if for every pair of paths α, β in $\mathcal{U}$ with the same end points, α and β are homotopic in X with fixed end points (cf. Corollary 9.9).
2. If $\mathcal{U} \in \mathcal{B}$ and $\mathcal{V} \subset \mathcal{U}$ is a path-connected open subset, then $\mathcal{V} \in \mathcal{B}$.
3. $\mathcal{B}$ is a base for the topology of X.

In particular, the third property holds because X is reasonable: every point $x \in X$ has a simply connected neighborhood, which contains an open neighborhood that necessarily belongs to $\mathcal{B}$, and it follows that every open subset of X is a union of such sets.

Now for any $\mathcal{U} \in \mathcal{B}$ with a point $x \in \mathcal{U}$ and a path γ in X from x_0 to x, let

$$\mathcal{U}_{[\gamma]} := \left\{ [\gamma \cdot \alpha] \in \widetilde{X} \;\middle|\; \alpha \text{ is a path in } \mathcal{U} \text{ starting at } x \right\}.$$

Notice that $\mathcal{U}_{[\gamma]}$ depends only on the homotopy class $[\gamma] \in \widetilde{X}$; this relies on the fact that since $\mathcal{U} \in \mathcal{B}$, the path α in the definition above is uniquely determined up to homotopy in X by its end point. It follows in fact that $p : \widetilde{X} \to X$ restricts to a bijection

$$\mathcal{U}_{[\gamma]} \xrightarrow{p} \mathcal{U}.$$

With all this in mind, one can now show that

$$\widetilde{\mathcal{B}} := \left\{ \mathcal{U}_{[\gamma]} \subset \widetilde{X} \;\middle|\; \mathcal{U} \in \mathcal{B} \text{ and } [\gamma] \in \widetilde{X} \text{ with } \gamma(1) \in \mathcal{U} \right\}$$

is a base for a topology on $\widetilde{X}$ such that each $\mathcal{U} \in \mathcal{B}$ is evenly covered by $p : \widetilde{X} \to X$. We leave the details of this as an exercise.

There is an obvious choice of base point in $\widetilde{X}$: define $\tilde{x}_0 \in \widetilde{X}$ as the homotopy class of the constant path at x_0. It remains to prove that $\pi_1(\widetilde{X}, \tilde{x}_0) = 0$. Since we now know that $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ is a covering map, Corollary 15.13 implies that $p_* : \pi_1(\widetilde{X}, \tilde{x}_0) \to \pi_1(X, x_0)$ is injective, thus it will suffice to show that the subgroup $p_* \pi_1(\widetilde{X}, \tilde{x}_0)$ in $\pi_1(X, x_0)$ is trivial. This subgroup is the set of homotopy classes $[\gamma] \in \pi_1(X, x_0)$ for which the loop γ lifts to a loop $\tilde{\gamma}$ based at $\tilde{x}_0$. The lift of γ to $\widetilde{X}$ can be written as

$$\tilde{\gamma}(t) = [\gamma_t] \in \widetilde{X},$$

where for each $t \in I$ we define

$$\gamma_t(s) := \begin{cases} \gamma(s) & \text{for } 0 \leqslant s \leqslant t, \\ \gamma(t) & \text{for } t \leqslant s \leqslant 1. \end{cases}$$

Then assuming $\tilde{\gamma}$ is a loop, we find $\tilde{\gamma}(1) = [\gamma] = \tilde{\gamma}(0) = [\text{const}]$, which is simply the statement that γ is homotopic with fixed end points to a constant loop, hence $[\gamma] \in \pi_1(X, x_0)$ is the trivial element. ☐

I do not have the energy to draw the picture myself, but I highly recommend looking at the picture of the universal cover of $S^1 \vee S^1$ on page 59 of [**Hat02**]. The idea here is that for every homotopically nontrivial loop in $S^1 \vee S^1$, one obtains a non-closed path in the universal cover $\widetilde{X}$. One can thus construct $\widetilde{X}$ one path at a time if one denotes by a and b the generators of $\pi_1(S^1 \vee S^1, x) \cong F_{\{a,b\}}$: at each step, the loops a, b, a^{-1} and b^{-1} furnish four homotopically distinct choices of loops to traverse, which lift to four distinct paths in $\widetilde{X}$ from one copy of the base point to another. Starting at the natural base point $\tilde{x}_0$ and following this procedure recursively produces the fractal picture in [**Hat02**, p. 59].

The application to the Galois correspondence requires a brief digression on topological groups and group actions.

Definition 17.3. A **topological group** (*topologische Gruppe*) is a group G with a topology such that the maps

$$G \times G \to G : (g, h) \mapsto gh \quad \text{and} \quad G \to G : g \mapsto g^{-1}$$

are both continuous.

Popular examples of topological groups include the various subgroups of the real or complex general linear groups $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(n, \mathbb{C})$, e.g. the orthogonal group $\mathrm{O}(n)$ and unitary group $\mathrm{U}(n)$, the special linear groups $\mathrm{SL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathbb{C})$, and so forth. We saw in Exercise 7.29 that for any locally compact and locally connected Hausdorff space X, the group of homeomorphisms $\mathrm{Homeo}(X)$ is a topological group with the group operation defined by composition. Finally, *any* group can be regarded as a topological group if we assign to it the discrete topology; this follows from the fact that every map on a space with the discrete topology is continuous. Topological groups with the discrete topology are often referred to as **discrete groups**.

Definition 17.4. Given a topological group G and a space X, a (continuous) G-action (*Wirkung*) on X is a (continuous) map

$$G \times X \to X : (g, x) \mapsto g \cdot x$$

such that the identity element $e \in G$ satisfies $e \cdot x = x$ for all $x \in X$ and $(gh) \cdot x = g \cdot (h \cdot x)$ holds for all $g, h \in G$ and $x \in X$.

Notice that for any G-action on X, there is a natural group homomorphism $G \to \mathrm{Homeo}(X)$ sending $g \in G$ to the homeomorphism $\varphi_g : X \to X$ defined by $\varphi_g(x) = g \cdot x$. If G is a discrete group then the converse is also true: every group homomorphism $G \to \mathrm{Homeo}(X)$ comes from a G-action on X. This is true because as long as the topology of G is discrete, the map $G \times X \to X : (g, x) \mapsto g \cdot x$ is continuous if and only if the map $X \to X : x \mapsto g \cdot x$ is continuous for every fixed $g \in G$. If G has a more interesting topology, then continuity of the map $(g, x) \mapsto g \cdot x$ with respect to $g \in G$ is also a nontrivial condition that would need to be checked—but we have no need to worry about this right now, as most of the groups we will deal with below are discrete.

Example 17.5. For any covering map $p : Y \to X$, $\mathrm{Aut}(p)$ acts as a discrete group on Y by $f \cdot y := f(y)$.

EXAMPLE 17.6. Regarding $\mathbb{Z}_2$ as a discrete group, a $\mathbb{Z}_2$-action on any space X is determined by the homeomorphism $\varphi_1 : X \to X$ associated to the nontrivial element $[1] \in \mathbb{Z}/2\mathbb{Z} =: \mathbb{Z}_2$, and this is necessarily an **involution**, i.e. it is its own inverse. A frequently occurring example is the action of $\mathbb{Z}_2$ on S^n defined via the antipodal map $\mathbf{x} \mapsto -\mathbf{x}$.

EXAMPLE 17.7. Here is a non-discrete example: any subgroup of the orthogonal group $\mathrm{O}(n)$ acts on $S^{n-1} \subset \mathbb{R}^n$ by matrix-vector multiplication, $A \cdot \mathbf{x} = A\mathbf{x}$.

For any G-action on X and a subset $\mathcal{U} \subset X$, we denote

$$g \cdot \mathcal{U} := \{g \cdot x \mid x \in \mathcal{U}\} \subset X.$$

Similarly, for each point $x \in X$, we define its **orbit** (*Bahn*) as the subset

$$G \cdot x := \{g \cdot x \mid g \in G\} \subset X.$$

One can easily check that for any two points $x, y \in X$, their orbits $G \cdot x$ and $G \cdot y$ are either identical or disjoint, thus there is an equivalence relation $\sim$ on X such that $x \sim y$ if and only if $G \cdot x = G \cdot y$. The quotient topological space defined by this equivalence relation is denoted by

$$X/G := X/\sim \; = \{\text{orbits } G \cdot x \subset X \mid x \in X\}.$$

EXAMPLE 17.8. The quotient $S^n/\mathbb{Z}_2$ arising from the action in Example 17.6 is $\mathbb{RP}^n$.

PROPOSITION 17.9. *Regarding $\pi_1(X, x_0)$ as a discrete group, any covering map $p : (Y, y_0) \to (X, x_0)$ of reasonable spaces with $\pi_1(Y) = 0$ gives rise to a natural action of $\pi_1(X, x_0)$ on Y.*

PROOF. There are at least two ways to see the action of $\pi_1(X, x_0)$ on a simply connected cover. First, Corollary 16.15 identifies $\pi_1(X, x_0)$ with $\mathrm{Aut}(p)$, and the latter acts on Y as explained in Example 17.5.

Alternatively, one can appeal to the uniqueness of the universal cover, so $p : (Y, y_0) \to (X, x_0)$ is necessarily isomorphic to the specific cover $\widetilde{X} = \{\text{paths } x_0 \rightsquigarrow x\}/\underset{h+}{\sim}$ that we constructed in the proof of Theorem 16.10. Then the obvious way for homotopy classes of loops $[\alpha] \in \pi_1(X, x_0)$ to act on homotopy classes of paths $[\gamma] \in \widetilde{X}$ is by concatenation:

$$[\alpha] \cdot [\gamma] := [\alpha \cdot \gamma].$$

It is easy to verify that this also defines a group action. □

EXERCISE 17.10. Show that the two actions of $\pi_1(X, x_0)$ on the universal cover constructed in the above proof are the same.

DEFINITION 17.11. A G-action on X is **free** (*frei*) if the only element $g \in G$ satisfying $g \cdot x = x$ for some $x \in X$ is the identity $g = e$.

The action is called **properly discontinuous** (*eigentlich diskontinuierlich*) if every $x \in X$ has a neighborhood $\mathcal{U} \subset X$ such that

$$(g \cdot \mathcal{U}) \cap \mathcal{U} = \varnothing$$

for every $g \in G$ with $g \cdot x \neq x$.

EXERCISE 17.12. Show that if a G-action is free and properly discontinuous, then G is discrete.

EXERCISE 17.13. Show that for any covering map $p : Y \to X$, the action of $\mathrm{Aut}(p)$ on Y as in Example 17.5 is free and properly discontinuous.

The observation that actions of deck transformation groups are free already has some nontrivial consequences, for instance:

PROPOSITION 17.14. *There exists no covering map $p : \mathbb{D}^2 \to X$ with $\deg(p) > 1$.*

Proof. If $\deg(p) > 1$, then since $\pi_1(\mathbb{D}^2) = 0$, we observe that the cover $p : \mathbb{D}^2 \to X$ must be regular and therefore has a nontrivial deck transformation group $\mathrm{Aut}(p)$ which acts freely on $\mathbb{D}^2$. But the Brouwer fixed point theorem rules out the existence of any nontrivial free group action on $\mathbb{D}^2$. ☐

The main purpose of the above definitions is that they lead to the following theorem, whose proof is now an easy exercise.

Theorem 17.15. *If G acts on X freely and properly discontinuously, then the quotient projection*

$$q : X \to X/G : x \mapsto G \cdot x$$

is a regular covering map with $\mathrm{Aut}(q) = G$. ☐

Now we are ready to finish the proof of the Galois correspondence.

Proof of Theorem 16.9. We have already shown that the correspondence is well defined and injective, so we need to prove surjectivity, in other words: given a reasonable space X with base point $x_0 \in X$ and any subgroup $H \subset G := \pi_1(X, x_0)$, we need to find a reasonable space Y with a covering map $p : (Y, y_0) \to (X, x_0)$ such that $p_*\pi_1(Y, y_0) = H$. Since X is reasonable, there exists a universal cover $f : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$, whose automorphism group is isomorphic to G, so this isomorphism defines a free and properly discontinuous action of G on $\widetilde{X}$. It also defines a free and properly discontinuous action of every subgroup of G on $\widetilde{X}$, and in particular an H-action. Define

$$Y := \widetilde{X}/H \quad \text{and} \quad p : Y \to X : H \cdot \tilde{x} \mapsto f(\tilde{x}).$$

It is straightforward to check that this is a covering map, and it is base-point preserving if we define $y_0 := H \cdot \tilde{x}_0$ as the base point of Y. Moreover, the quotient projection $q : (\widetilde{X}, \tilde{x}_0) \to (Y, y_0)$ is now the universal cover of Y, and it fits into the following commutative diagram:

$$\begin{tikzcd}
(\widetilde{X}, \tilde{x}_0) \arrow[r, "f"] \arrow[d, "q"] & (X, x_0) \\
(Y, y_0) \arrow[ur, "p"] &
\end{tikzcd}$$

Given a loop γ in X based at x_0, let γ' denote its lift to a path in Y starting at y_0, and let $\tilde{\gamma}$ denote the lift to a path in $\widetilde{X}$ starting at $\tilde{x}_0$, The subgroup $p_*\pi_1(Y, y_0) \subset \pi_1(X, x_0)$ is precisely the set of all homotopy classes $[\gamma] \in \pi_1(X, x_0)$ for which γ' is a loop. Notice that since all maps in the diagram are covering maps, $\tilde{\gamma}$ is also a lift of γ' via the covering map q. Then $[\gamma] \in H$ so that γ' is a loop if and only if the end point of $\tilde{\gamma}$ is in $q^{-1}(y_0) = H \cdot \tilde{x}_0$. Under the natural bijection between $\pi_1(X, x_0)$ and $f^{-1}(x_0) = G \cdot \tilde{x}_0$, this just means $[\gamma] \in H$, hence $p_*\pi_1(Y, y_0) = H$. ☐

18. Manifolds

I have mentioned manifolds already a few times in this course, but now it is time to discuss them somewhat more precisely. While we do not plan to go to deeply into this subject this semester, the goal is in part to understand what the main definitions are and why, forming the basis of the subject known as "geometric topology". In so doing, we will also establish an inventory of examples and concepts that will serve as useful intuition when we start to talk about homology next week.

Definition 18.1. A **topological manifold** (*Mannigfaltigkeit*) of dimension $n \geqslant 0$ (often abbreviated with the term "n-manifold") is a second countable Hausdorff space M such that every point $p \in M$ has a neighborhood homeomorphic to $\mathbb{R}^n$.

More generally, a **topological n-manifold with boundary** (*Mannigfaltigkeit mit Rand*) is a second countable Hausdorff space M such that every point $p \in M$ has a neighborhood homeomorphic to either $\mathbb{R}^n$ or the so-called "n-dimensional half-space"

$$\mathbb{H}^n := [0, \infty) \times \mathbb{R}^{n-1}.$$

The third condition in each of these definitions is probably the most intuitive and is the most distinguishing feature of manifolds: we abbreviate it by saying that manifolds are "locally Euclidean". It means in effect that sufficiently small open subsets of a manifold can be described via *local coordinate systems.* The technical term for this is "chart": a **chart** (*Karte*) on an n-manifold with boundary is a homeomorphism

$$\varphi : \mathcal{U} \to \Omega$$

where $\mathcal{U} \subset M$ and $\Omega \subset \mathbb{H}^n$ are open subsets. As special cases, Ω may be the whole of $\mathbb{H}^n$, or an open ball in $\mathbb{H}^n$ disjoint from

$$\partial\mathbb{H}^n := \{0\} \times \mathbb{R}^{n-1},$$

in which case Ω is also homeomorphic to $\mathbb{R}^n$. It follows that on any n-manifold (with or without boundary), every point is in the domain of a chart. Conversely, if we are given a collection of charts $\{\varphi_\alpha : \mathcal{U}_\alpha \to \Omega_\alpha\}_{\alpha \in J}$ such that $M = \bigcup_{\alpha \in J} \mathcal{U}_\alpha$, then after shrinking the domains and targets of these charts if necessary, we can assume every point $p \in M$ is in the domain of some chart $\varphi_\alpha : \mathcal{U}_\alpha \to \Omega_\alpha$ such that Ω_α is either an open ball in $\mathbb{H}^n \backslash \partial\mathbb{H}^n$ or a half-ball with boundary on $\partial\mathbb{H}^n$, so that Ω is homeomorphic to either $\mathbb{R}^n$ or $\mathbb{H}^n$. This means M is locally Euclidean, so both versions of the third condition in our definition can be rephrased as the condition that M is covered by charts. The **boundary** of a manifold M with boundary can now be defined as the subset

$$\partial M := \left\{ p \in M \;\middle|\; \varphi(p) \in \partial\mathbb{H}^n \text{ for some chart } \varphi \right\},$$

which is clearly an $(n-1)$-manifold (without boundary).

The word "topological" is included before "manifold" in order to make the distinction between topological manifolds and *smooth manifolds*, which we will discuss a little bit below. By default in this course, you should assume that everything we refer to simply as a "manifold" is actually a *topological* manifold unless otherwise specified. (If this were a differential geometry course, you would instead want to assume that "manifold" always means *smooth* manifold.) One can regard manifolds without boundary as being special cases of manifolds M with boundary such that $\partial M = \varnothing$, so we shall also use "manifold" as an abbreviation for the term "manifold with boundary" and will generally specify "without boundary" when we want to assume $\partial M = \varnothing$. You should be aware that some books adopt different conventions for such details, e.g. some authors assume $\partial M = \varnothing$ always unless the words "with boundary" are explicitly included.

REMARK 18.2. The following detail deserves emphasis: the way we have expressed the definition of the boundary $\partial M \subset M$ above makes sense in part because when we defined the notion of a chart $\varphi : \mathcal{U} \to \Omega$, we required[24] its image Ω to be an open subset of the half-space $\mathbb{H}^n$, and not necessarily an open subset of $\mathbb{R}^n$. If we were allowing arbitrary open subsets $\Omega \subset \mathbb{R}^n$, then every point $p \in M$ would be a boundary point, because e.g. one could take any chart $\varphi : \mathcal{U} \to \Omega$ with $p \in \mathcal{U}$ and compose it with a translation on $\mathbb{R}^n$ so that $\varphi(p) = 0 \in \partial\mathbb{H}^n$. Requiring $\Omega \subset \mathbb{H}^n$ prevents this in general, because if we start with a chart $\varphi : \mathcal{U} \to \Omega$ whose image contains an open ball around $\varphi(p)$, then translating it to achieve $\varphi(p) = 0$ will produce something whose image cannot be contained in $\mathbb{H}^n$. In fact, the translation trick works only for points $p \in \mathcal{U}$ with $\varphi(p) \in \partial\mathbb{H}^n$, as

---

[24] This convention is not universal: many books allow charts to have images that are arbitrary open subsets of $\mathbb{R}^n$. The latter is a sensible convention especially if one only wants to consider manifolds with empty boundary, and even if nonempty boundaries are allowed, one can work with charts defined in this way, but the definition of $\partial M \subset M$ would need to be expressed a bit differently.

these are precisely the points for which Ω does not contain any ball around $\varphi(p)$. It can happen that $\Omega \subset \mathbb{H}^n$ is *also* an open subset of $\mathbb{R}^n$: this is true if and only if $\Omega \cap \partial\mathbb{H}^n = \varnothing$, and in that case, none of the points in the domain of the chart are boundary points. One can show that whenever $\varphi(p) \in \partial\mathbb{H}^n$ for some chart $\varphi : \mathcal{U} \to \Omega$ with $p \in \mathcal{U}$, the same must hold for all other charts whose domains contain p; in other words, no point of M can be simultaneously a boundary point and an *interior* point, where the latter means that some chart maps it into $\mathbb{H}^n \backslash \partial\mathbb{H}^n$. For $n \leqslant 2$, this can be proved using methods that we have already developed (see Exercise 19.13); the proof for $n > 2$ requires some other methods that we haven't developed yet, but will soon, e.g. singular homology.

Manifolds are usually what we have in mind when we think of spaces that are "nice" or "reasonable". In particular, the following is an immediate consequence of the observation that every point in $\mathbb{R}^n$ or $\mathbb{H}^n$ has a neighborhood homeomorphic to the closed n-disk:

PROPOSITION 18.3. *For an n-manifold M and a point $p \in M$, every neighborhood of p contains one that is homeomorphic to $\mathbb{D}^n$.* ☐

COROLLARY 18.4. *Manifolds are locally compact and locally path-connected. They are also* ***locally contractible****, meaning every neighborhood of every point in M contains a contractible neighborhood. In particular, they are "reasonable" in the sense of Definition 16.1.* ☐

It follows via Theorem 7.19 that a manifold M is connected if and only if it is path-connected. More generally, the path-components of M are the same as its connected components (cf. Prop. 7.18), each of which are open and closed subsets, hence M is homeomorphic to the disjoint union of its connected components. It is similarly easy to show that these connected components are also manifolds.

DEFINITION 18.5. A manifold M is **closed** (*geschlossen*) if it is compact and $\partial M = \varnothing$. It is **open** (*offen*) if none of its connected components are closed, i.e. all of them either are noncompact or have nonempty boundary.

You need to be aware that these usages of the words "closed" and "open" are different from the notions of closed or open subsets in a topological space. The distinction between a "closed manifold" and a "closed subset" is at least more explicit in German: the former is a *geschlossene Mannigfaltigkeit*, while the latter is an *abgeschlossene Teilmenge*. For openness there is the same ambiguity in German and English, but it is rarely a problem: you just need to pay attention to the context in which these adjectives are used and what kinds of nouns they are modifying. We will not have much occasion to talk about open manifolds in this course, and many authors apparently dislike seeing the word "open" used in this way, but it has some advantages, e.g. in differential topology, there are some elegant theorems that can be stated most naturally for open manifolds but are not true for manifolds that are not open.

EXAMPLE 18.6. Any discrete space with only countably many points is a 0-manifold. (Discrete spaces with uncountably many points are excluded because they are not second countable.) Conversely, this is an accurate description of every 0-manifold, and the closed ones are those that are finite. Note that a 0-manifold can never have boundary.

EXAMPLE 18.7. The line $\mathbb{R}$, the interval $(-1, 1)$ and the circle S^1 are all examples of 1-manifolds without boundary, where S^1 is closed and the others are open. Further examples without boundary are obtained by taking arbitrary countable disjoint unions of these examples, e.g. $S^1 \amalg \mathbb{R}$ is a 1-manifold without boundary, though it is neither closed nor open since it has one closed component and one that is not closed. Some examples of 1-manifolds with nonempty boundary include the interval $I = [0, 1]$, whose boundary is the compact 0-manifold $\partial I = \{0, 1\}$, and $[0, 1)$, whose boundary is $\partial[0, 1) = \{0\}$.

Example 18.8. The word **surface** (*Fläche*) refers in general to a 2-dimensional manifold. Examples without boundary include S^2, $\mathbb{T}^2 = S^1 \times S^1$, the surfaces Σ_g of genus $g \geqslant 0$, $\mathbb{RP}^2$, $\mathbb{R}^2$, and arbitrary countable disjoint unions of any of these. One can also take connected sums of these examples to obtain more, though as we've seen, not all of the examples that arise in this way are new, e.g. Σ_g for $g \geqslant 1$ is the g-fold connected sum of copies of $\mathbb{T}^2$. Some compact examples with boundary include $\mathbb{D}^2$ (with $\partial\mathbb{D}^2 = S^1$) and the surface $\Sigma_{g,m}$ of genus g with $m \geqslant 1$ holes cut out, which has $\partial\Sigma_{g,m} \cong \coprod_{i=1}^m S^1$. An obvious noncompact example with nonempty boundary is the half-plane $\mathbb{H}^2$, with $\partial\mathbb{H}^2 \cong \mathbb{R}$.

Example 18.9. Some examples of arbitrary dimension n without boundary are S^n, $\mathbb{RP}^n$, $\mathbb{R}^n$, $\mathbb{T}^n := S^1 \times \ldots \times S^1$, any open subset of any of these, and anything obtained from these by (countable) disjoint unions or connected sums.[25] Some obvious examples with nonempty boundary are $\mathbb{D}^n$ (with $\partial\mathbb{D}^n = S^{n-1}$), and $[-1,1] \times \mathbb{T}^{n-1}$, whose boundary is the disjoint union of two copies of $\mathbb{T}^{n-1}$.

While we don't plan to do very much with it in this course, we now make a brief digression on the subject of *smooth* manifolds, which are the main object of study in differential geometry and differential topology. As preparation, observe that if $\varphi_\alpha : \mathcal{U}_\alpha \to \Omega_\alpha$ and $\varphi_\beta : \mathcal{U}_\beta \to \Omega_\beta$ are two charts on the same manifold M, then on any region $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ where they overlap, we can think of them as describing two alternative coordinate systems, so that there is a well-defined "coordinate transformation" map switching from one to the other. To be more precise, $\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ and $\varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ are open subsets of Ω_α and Ω_β respectively, and there is a homeomorphism from one to the other defined via the following diagram:

$$\begin{tikzcd}
& \mathcal{U}_\alpha \cap \mathcal{U}_\beta \arrow[dl, "\varphi_\alpha"'] \arrow[dr, "\varphi_\beta"] & \\
\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \arrow[rr, "\varphi_\beta \circ \varphi_\alpha^{-1}"] & & \varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)
\end{tikzcd}$$

The map $\varphi_\beta \circ \varphi_\alpha^{-1}$ is called the **transition map** (*Übergang*) relating φ_α and φ_β. The key point about a transition map is that its domain and target are open subsets of a Euclidean space (or half-space), thus we know what it means for such a map to be "differentiable". This observation makes it possible to do differential calculus on manifolds and to speak of functions $f : M \to \mathbb{R}$ as being differentiable or not: the idea is that f should be called differentiable if it appears differentiable whenever it is written in a local coordinate system. But for this to be well defined, we need to be assured that the answer to the differentiability question will not change if we change coordinate systems, i.e. if we compose our local coordinate expression for f with a transition map. If all conceivable charts for M are allowed, then the answer will indeed sometimes change, because the composition of a differentiable function with a non-differentiable map is not usually differentiable. We therefore need to be able to assume that transition maps are always differentiable, and since this is not true if all conceivable charts are allowed, we need to restrict the class of charts that we consider. This restriction introduces a bit of structure on M that is not determined by its topology, but is something extra:

Definition 18.10. A **smooth structure** (*glatte Struktur*) on an n-dimensional topological manifold M is a maximal collection of charts $\{\varphi_\alpha : \mathcal{U}_\alpha \to \Omega_\alpha\}_{\alpha \in J}$ for which $M = \bigcup_{\alpha \in J} \mathcal{U}_\alpha$ and the corresponding transition maps $\varphi_\beta \circ \varphi_\alpha^{-1}$ for all $\alpha, \beta \in J$ are of class C^∞. A topological manifold endowed with a smooth structure is called a **smooth manifold** (*glatte Mannigfaltigkeit*).

---

[25] Recall from Lecture 13 the connected sum of two n-manifolds M and N: it is defined by deleting the interiors of two embedded n-disks from M and N and then gluing them together along the spheres S^{n-1} at the boundaries of these disks.

It is easy to see that a single topological manifold can have multiple distinct smooth structures, e.g. on $M = \mathbb{R}$, the functions $\varphi_\alpha(t) = t$ and $\varphi_\beta(t) = t^3$ are homeomorphisms $\mathbb{R} \to \mathbb{R}$ and can thus be regarded as charts, but $\varphi_\alpha \circ \varphi_\beta^{-1}$ is not everywhere differentiable, hence φ_α and φ_β can each be regarded as belonging to smooth structures on $\mathbb{R}$, but they are distinct smooth structures. That is a relatively uninteresting example, but there are also known examples of topological manifolds admitting multiple smooth structures that are not even equivalent up to *diffeomorphism* (the smooth version of homeomorphism), as well as topological manifolds that do not admit any smooth structure at all. Such things are very hard to prove, but you should not worry about them right now, because the basic fact is that most manifolds we encounter in nature have natural smooth structures. A very high proportion of them come from the following geometric version of the implicit function theorem.

THEOREM 18.11 (implicit function theorem). *Suppose $\mathcal{U} \subset \mathbb{R}^n$ is an open subset, $F : \mathcal{U} \to \mathbb{R}^k$ is a C^∞-map and $q \in \mathbb{R}^k$ is a point such that for all $p \in F^{-1}(q)$, the derivative $dF(p) : \mathbb{R}^n \to \mathbb{R}^k$ is surjective (we say in this case that q is a **regular value** of F). Then $F^{-1}(q) \subset \mathbb{R}^n$ is a smooth manifold of dimension $n - k$.* ☐

The above theorem is provided “for your information,” meaning we do not plan to either prove or use it in any serious way in this course, but you should be aware that it exists because it provides many examples of manifolds that arise naturally in various applications. For instance:

EXAMPLE 18.12. The n-sphere $S^n = F^{-1}(1)$, where $F : \mathbb{R}^{n+1} \to \mathbb{R} : \mathbf{x} \mapsto |\mathbf{x}|^2$, which has 1 as a regular value.

EXAMPLE 18.13. The special linear group $\mathrm{SL}(n, \mathbb{R}) = \det^{-1}(1)$ for the determinant map $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$. One can show that 1 is a regular value of det by relating the derivative of the determinants of a family of matrices passing through $\mathbb{1}$ to the trace of the derivative of that family of matrices. Thus $\mathrm{SL}(n, \mathbb{R})$ is a smooth manifold of dimension $n^2 - 1$.

Now let’s look at a couple of non-examples.

EXAMPLE 18.14. The wedge sum $S^1 \vee S^1$ is *not* a manifold of any dimension. It does look like a 1-manifold in the complement of the base point $x \in S^1 \vee S^1$, but x does not have any neighborhood homeomorphic to Euclidean space. Indeed, sufficiently small neighborhoods $\mathcal{U} \subset S^1 \vee S^1$ of x all look like two line segments intersecting, so that if we delete the point x, we obtain a space $\mathcal{U}\backslash\{x\}$ with four path-components. This cannot happen in an n-manifold for any n, as deleting a point from $\mathbb{R}$ produces two path-components, while deleting a point from $\mathbb{R}^n$ with $n \geqslant 2$ leaves a space that is still path-connected.

EXAMPLE 18.15. Here is a space that is locally Euclidean and second countable, but not Hausdorff: the line with two zeroes, i.e. $X := (\mathbb{R} \times \{0, 1\})/\sim$ with $(x, 0) \sim (x, 1)$ for all $x \neq 0$. If we endow X with the quotient topology induced by the natural topology of $\mathbb{R} \times \{0, 1\} \cong \mathbb{R} \amalg \mathbb{R}$, then a subset $\mathcal{U} \subset X$ is open if and only if its preimage under the quotient projection $\mathbb{R} \times \{0, 1\} \to X$ is open, and it follows in particular that the images of $\mathbb{R} \times \{0\}$ and $\mathbb{R} \times \{1\}$ under this projection are open subsets of X that are each (in obvious ways) homeomorphic to $\mathbb{R}$. The two zeroes $0_0 := [(0, 0)]$ and $0_1 := [(0, 1)]$ therefore each have neighborhoods homeomorphic to $\mathbb{R}$, and so (for more obvious reasons) does every other point, so the line with two zeroes would count as a 1-manifold if we did not require manifolds to be Hausdorff. We should emphasize that we are considering the quotient topology on X, not the pseudometric topology (cf. Example 6.12); X with the pseudometric topology is not locally homeomorphic to $\mathbb{R}$, because every neighborhood of 0_0 must also contain 0_1 and vice versa, so the two subsets described above would no longer be open.

Example 18.16. The following is a compact variation on the previous example: writing X for the line with two zeroes, its one point compactification X^* is obtained by adding a single point called ∞, which is the limit of any sequence in X that has no bounded subsequence. Just as the one point compactification $\mathbb{R} \cup \{\infty\}$ of $\mathbb{R}$ is homeomorphic to S^1, we can think of X^* as the result of replacing one point $0 \in \mathbb{R} \subset S^1$ with a pair of points $0_0, 0_1 \in X^*$ that each have neighborhoods homeomorphic to $\mathbb{R}$, but with every neighborhood of 0_0 intersecting every neighborhood of 0_1. This would also be a 1-manifold if manifolds were not required to be Hausdorff.

You probably don't need much convincing by this point that spaces which are Hausdorff and second countable are "good," while those that lack either of these properties are "bad". Nonetheless, it's worth taking a moment to consider *why* it would be bad if we dropped either of these conditions from the definition of a manifold. The first answer is clearly that if we dropped the Hausdorff axiom, then Example 18.15 would be a manifold, and we don't like Example 18.15. But there are better reasons. One of them is related to the implicit function theorem, Theorem 18.11 above, which produces many examples of manifolds that are subsets of larger-dimensional Euclidean spaces. Notice that in this situation, it is completely unnecessary to verify whether those subsets are Hausdorff or second countable, because every subset of a finite-dimensional Euclidean space is both. (See Exercise 5.9 if you've forgotten how we know that $\mathbb{R}^n$ is second countable.) Now, it is reasonable to ask whether *all* conceivable manifolds arise from something similar to Theorem 18.11, i.e. are all of them embeddable into $\mathbb{R}^N$ for some $N \in \mathbb{N}$? The answer is yes, though clearly it would not be if the Hausdorff and second countability conditions were not included:

Theorem 18.17. *Every topological manifold is homeomorphic to a closed subset of* $\mathbb{R}^N$ *for* $N \in \mathbb{N}$ *sufficiently large.* ☐

This is another theorem that I am providing "for your information," as I do not intend to use it for anything and therefore will not prove it. A readable proof for the case of a compact manifold appears in [**Hat02**, Corollary A.9]. The noncompact case is significantly harder and proofs typically do not appear in textbooks, but the idea is outlined and some precise references given in [**Lee11**, p. 116]. I would caution you in any case against taking this theorem more seriously than it deserves: while it's nice to know that all manifolds are in some sense *submanifolds* of some $\mathbb{R}^N$, many of them do not come with any canonical choice of embedding into $\mathbb{R}^N$, so this property is not in any way intrinsic to their structure and one should (and usually can) avoid using it to prove things about manifolds. It might also be argued that Theorem 18.17 undermines my point about the Hausdorff and second countability assumptions being indispensable, since it may seem desirable to be able to consider "manifolds" that are *more general* than just submanifolds of Euclidean spaces.

As a general principle, mathematicians consider a definition to be a "good" definition if it appears as the hypothesis for a good theorem. I'm not sure if Theorem 18.17 truly qualifies as a good theorem. But I want to talk about another one that I think is better.

Theorem 18.18. *Every connected nonempty* 1*-manifold without boundary is homeomorphic to either* S^1 *or* $\mathbb{R}$.

If this statement sounds at first too restrictive, it makes up for it by being extremely useful. In combination with the implicit function theorem, one can deduce from it e.g. the possible topologies of regular level sets of arbitrary smooth functions $F : \mathbb{R}^n \to \mathbb{R}^{n-1}$. This ability has a surprising number of beautiful applications in differential topology and related fields; one example is the definition of the "mapping degree," sketched in Exercise 19.14. Those applications are typically based on the following corollary for compact manifolds with boundary.

Corollary 18.19. *Every compact* 1*-manifold* M *with boundary is homeomorphic to a disjoint union of finitely many copies of* S^1 *and* $[0,1]$*. In particular,* ∂M *consists of evenly many points.*

Proof. Since M is compact, it can have at most finitely many connected components (otherwise we can find a noncompact closed subset by choosing one point from every component). Restricting to connected components, it will therefore suffice to show that every connected compact 1-manifold M is either S^1 or $[0,1]$. Theorem 18.18 implies that $M \cong S^1$ if $\partial M = \varnothing$, so assume otherwise. Then ∂M is a closed subset and therefore is compact, and it is also a 0-manifold, which means it is a nonempty finite set. Let us modify M by attaching a half-line $[0,\infty)$ to each boundary point, that is, let

$$\widehat{M} := M \cup_{\partial M} \left(\coprod_{p \in \partial M} [0,\infty) \right).$$

This makes $\widehat{M}$ a noncompact connected 1-manifold with empty boundary, so by Theorem 18.18, $\widehat{M} \cong \mathbb{R}$. It follows that $M \subset \widehat{M}$ is homeomorphic to a path-connected compact subset of $\mathbb{R}$. All such subsets are compact intervals $[a,b]$, hence $M \cong [0,1]$. □

The proof of Theorem 18.18 given below is based on a series of exercises outlined in [**Gal87**]. I will not go through every step in exhaustive detail, as my main objective is just to point out explicitly where the Hausdorff and second countability conditions are needed. You saw already from Examples 18.15 and 18.16 that the theorem becomes false if the Hausdorff condition is dropped, and after the proof we will look at an even stranger example to see what can happen without second countability.

Here is a lemma that depends explicitly on the Hausdorff property, e.g. you will find if you look again at the line with two zeroes (Example 18.15) that it is not satisfied in that particular example.

Lemma 18.20. *Suppose M is a Hausdorff space with two overlapping open subsets $\mathcal{U}_\alpha, \mathcal{U}_\beta \subset M$ that are each homeomorphic to $\mathbb{R}$, and neither is contained in the other. Then each connected component $\mathcal{W}$ of $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ is homeomorphic to $\mathbb{R}$ and has compact closure $\overline{\mathcal{W}} \subset M$ homeomorphic to $[0,1]$, whose boundary consists of a point $p_\alpha \in \mathcal{U}_\alpha$ that is not in $\mathcal{U}_\beta$ and a point $p_\beta \in \mathcal{U}_\beta$ that is not in $\mathcal{U}_\alpha$.*

Proof. Choose explicit homeomorphisms $\varphi_\alpha : \mathcal{U}_\alpha \to \mathbb{R}$ and $\varphi_\beta : \mathcal{U}_\beta \to \mathbb{R}$. The image $\varphi_\beta(\mathcal{W}) \subset \mathbb{R}$ is necesarily a connected open subset of $\mathbb{R}$, and is therefore an open interval, implying $\mathcal{W} \cong \mathbb{R}$. But $\varphi_\beta(\mathcal{W})$ cannot be the entirety of $\mathbb{R}$, as that would imply $\mathcal{W} = \mathcal{U}_\beta$ since φ_β is a homeomorphism, and thus $\mathcal{U}_\beta \subset \mathcal{U}_\alpha$, which was excluded in the hypotheses. For the same reasons, $\varphi_\alpha(\mathcal{W})$ is an open interval in $\mathbb{R}$, but not the entirety of $\mathbb{R}$.

Let us show that the closure $\overline{\mathcal{W}} \subset M$ contains two boundary points p_α, p_β with the stated properties. To find p_α, choose a point $t \in \mathbb{R}$ that is in the closure of $\varphi_\alpha(\mathcal{W}) \subset \mathbb{R}$ but not in $\varphi_\alpha(\mathcal{W})$. Since φ_α is a homeomorphism, there must then exist a sequence $x_n \in \mathcal{W}$ converging to a point $p_\alpha := \varphi_\alpha^{-1}(t) \in \mathcal{U}_\alpha$, and p_α cannot belong to $\mathcal{U}_\beta$ since this would imply $p_\alpha \in \mathcal{W}$ and thus $t \in \varphi_\alpha(\mathcal{W})$. We claim: $|\varphi_\beta(x_n)| \to \infty$. Indeed, if this does not hold, then after replacing x_n with a suitable subsequence, we can assume $\varphi_\beta(x_n)$ converges to some point $y \in \mathbb{R}$, in which case x_n also converges to $x := \varphi_\beta^{-1}(y) \in \mathcal{U}_\beta$ since φ_β is a homeomorphism. But we already know $x_n \to p_\alpha$, so the assumption that M is Hausdorff implies $x = p_\alpha$ and gives a contradiction, since $p_\alpha \notin \mathcal{U}_\beta$.

It follows from the claim above that $\varphi_\beta(\mathcal{W}) \subset \mathbb{R}$ is an unbounded interval, and since it is not the entirety of $\mathbb{R}$, it is therefore an infinite half-interval of the form $(-\infty, a)$ or (b, ∞) for some $a, b \in \mathbb{R}$. Reversing the roles of α and β, a similar conclusion holds for $\varphi_\alpha(\mathcal{W})$, so for concreteness, let us suppose

$$\varphi_\alpha(\mathcal{W}) = (-\infty, a) \qquad \text{and} \qquad \varphi_\beta(\mathcal{W}) = (b, \infty),$$

in which case the recipe described above for defining $p_\alpha, p_\beta \in \overline{\mathcal{W}}$ gives

$$p_\alpha = \varphi_\alpha^{-1}(a), \qquad p_\beta = \varphi_\beta^{-1}(b).$$

(Only minor modifications to this discussion are necessary if $\varphi_\alpha(\mathcal{W})$ is instead bounded below or $\varphi_\beta(\mathcal{W})$ bounded above.) Moreover, the transition map

$$\mathbb{R} \supset \varphi_\alpha(\mathcal{W}) = (-\infty, a) \xrightarrow{\varphi_\beta \circ \varphi_\alpha^{-1}} (b, \infty) = \varphi_\beta(\mathcal{W}) \subset \mathbb{R},$$

being a homeomorphism between two open intervals in $\mathbb{R}$, is a monotone function whose value approaches $\pm\infty$ at the bounded end of its domain, and the same applies to its inverse, implying that this transition map also has a *finite* limit at the unbounded end of its domain. Now if $x_n \in \mathcal{W}$ is any sequence that has no subsequence converging to any point in $\mathcal{W}$ or to p_β, it follows that $|\varphi_\beta(x_n)| \to \infty$ and thus $\varphi_\alpha(x_n) \to a$, implying $x_n \to p_\alpha$. This proves that the union of $\mathcal{W}$ with the two points p_α, p_β is compact, as claimed. Putting the obvious topology on the extended interval $[b, \infty]$, φ_β now has a unique extension to a homeomorphism $\overline{\mathcal{W}} \to [b, \infty]$ that sends $p_\alpha \mapsto \infty$, so $\overline{\mathcal{W}}$ has the topology of a compact interval. ☐

Note that in the setting of the lemma, $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ may in general have multiple connected components, but the proof showed that a homeomorphism $\varphi_\alpha : \mathcal{U}_\alpha \to \mathbb{R}$ sends each of them to an unbounded half-interval. Here's a useful fact we know about $\mathbb{R}$: you can't fit more than two disjoint unbounded half-intervals into it!

COROLLARY 18.21. *In the setting of Lemma 18.20, $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ has either one or two connected components.* ☐

EXERCISE 18.22. Show that the compact non-Hausdorff space in Example 18.16 admits an open covering by two sets homeomorphic to $\mathbb{R}$ whose intersection with each other has three connected components.

PROOF OF THEOREM 18.18. Given a nonempty connected 1-manifold M without boundary, every point has an open neighborhood homeomorphic to $\mathbb{R}$, and since M is second countable, we can cover M with a *finite or countable* collection $\{\mathcal{U}_n \subset M\}_{n=1}^N$ of such neighborhoods with homeomorphisms $\varphi_n : \mathcal{U}_n \to \mathbb{R}$; here N is either a natural number or ∞. After removing some of these sets from the collection, we can assume without loss of generality that none of them are contained in any one of the others.

If $N = 1$, then M is homeomorphic to $\mathbb{R}$, and we are done.

If $N \geqslant 2$, then since M is also Hausdorff and connected, we can appeal to Lemma 18.20 and Corollary 18.21 in order to relabel the subsets $\{\mathcal{U}_n\}_{n=1}^N$ in the following manner. Choose $\mathcal{U}_1$ to be an arbitrary set in the collection. By definition $\mathcal{U}_1$ is an open subset of M, but it might also be a closed subset—if it is, then since M is connected, we can conclude that $M = \mathcal{U}_1 \cong \mathbb{R}$, so again we are done. If however $\mathcal{U}_1 \subset M$ is not a closed subset, then it is not the complement of any open set, and in particular it is not the complement of the union of the rest of the sets in our collection, which means at least one of them—which we shall now call $\mathcal{U}_2$—must intersect $\mathcal{U}_1$. There are now three possibilities:

1. If $\mathcal{U}_1 \cap \mathcal{U}_2$ has two connected components, one can deduce from Lemma 18.20 that $\mathcal{U}_1 \cup \mathcal{U}_2$ is homeomorphic to S^1, which is compact and is therefore (since M is Hausdorff) a closed subset of M. Since it is clearly also an open subset and M is connected, this implies $M = \mathcal{U}_1 \cup \mathcal{U}_2 \cong S^1$, so we are done.
2. If $\mathcal{U}_1 \cap \mathcal{U}_2$ has only one connected component, then $\mathcal{U}_1 \cup \mathcal{U}_2$ must be homeomorphic to $\mathbb{R}$. If $\mathcal{U}_1 \cup \mathcal{U}_2$ is also a closed subset of M, then connectedness again implies $M = \mathcal{U}_1 \cup \mathcal{U}_2 \cong \mathbb{R}$, and we are done.

(3) If $\mathcal{U}_1 \cap \mathcal{U}_2$ has only one connected component and the subset $\mathcal{U}_1 \cup \mathcal{U}_2 \subset M$ is not closed, then appealing again to the fact that M is connected, $\mathcal{U}_1 \cup \mathcal{U}_2$ must intersect one of the remaining subsets in our collection, which we shall now call $\mathcal{U}_3$.

Now repeat the previous step like so: if $(\mathcal{U}_1 \cup \mathcal{U}_2) \cap \mathcal{U}_3$ has two connected components, we can conclude $M = \mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{U}_3 \cong S^1$, and if not, then $\mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{U}_3 \cong \mathbb{R}$ and either this is all of M or it has nonempty intersection with one of the remaining sets in the collection. If the latter happens, repeat. And so on.

If N is finite, this process eventually exhausts all the sets $\mathcal{U}_1, \dots, \mathcal{U}_N$ and produces a homeomorphism of M to either S^1 or $\mathbb{R}$, the former if an intersection with two connected components ever occurs, and the latter otherwise.

If N is infinite, the process may still terminate if an intersection with two connected components appears, implying that finitely many of the sets $\mathcal{U}_n$ cover M and it is homeomorphic to S^1.

The remaining possibility is that the process never terminates, but instead produces a countable sequence of nested open subsets

$$I_1 \subset I_2 \subset I_3 \subset \dots \bigcup_{n=1}^{\infty} I_n = M,$$

where each $I_n := \mathcal{U}_1 \cup \dots \cup \mathcal{U}_n$ is homeomorphic to $\mathbb{R}$ and is obtained from I_{n-1} by gluing two copies of $\mathbb{R}$ together along a pair of connected half-intervals of infinite length. Up to homeomorphism, we could instead describe this process as follows: identify I_1 with $(0,1)$, and by induction, if I_{n-1} for some $n \geqslant 2$ has been identified with a finite interval (a,b), then I_n is identified with the union of (a,b) and another finite open interval that contains either a or b in its interior and has an end point in (a,b). Up to homeomorphism, we can thus assume $I_{n-1} = (a,b)$ and I_n is either $(a-1,b)$ or $(a,b+1)$. Continuing this process indefinitely, the union $\bigcup_{n=1}^{\infty} I_n$ gets identified with some subinterval in $\mathbb{R}$, and is thus homeomorphic to $\mathbb{R}$. □

The second countability axiom became relevant in the last step of this proof because M was presented as the union of a *countable* collection of intervals; if we had been forced to assume that the collection of Euclidean neighborhoods covering M was uncountable, we would not have been able to conclude in the same manner that M is homeomorphic to $\mathbb{R}$. I would now like to describe an example showing that this danger is serious, and that something other than S^1 or $\mathbb{R}$ can indeed arise if the second countability axiom is dropped. We will need to appeal to a rather non-obvious result from elementary set theory. Recall that a **totally ordered set** $(I, \prec)$ consists of a set I with a partial order $\prec$ such that for all pairs of elements $x, y \in I$, at least one of the conditions $x \prec y$ or $y \prec x$ holds. Such a set is said to be **well ordered** if every subset of I contains a smallest element. The most familiar example of a well-ordered set is the natural numbers. For the purposes of our example below, we need a well-ordered set that is uncountable.

LEMMA 18.23. *There exists an uncountable well-ordered set $(\omega_1, \leqslant)$ such that for every $x \in \omega_1$, at most countably many elements $y \in \omega_1$ satisfy $y \leqslant x$.*

Understanding this lemma requires some knowledge of the **ordinal numbers** (*Ordinalzahlen*), which we do not have time to describe here in detail, but the intuitive idea is to think of any well-ordered set as a "number," call two such numbers equivalent if there exists an order-preserving bijection from one to the other, and write $x \leqslant y$ whenever there exists an order-preserving injection from x into y. Informally, an ordinal number can be regarded as an equivalence class of well-ordered sets under this notion of equivalence. We can then think of each natural number $n \in \mathbb{N}$ as an ordinal number by identifying it with the set $\{1, \dots, n\}$, and this identification obviously produces the correct ordering relation for the natural numbers. But there are also infinite ordinal numbers, e.g. the set $\mathbb{N}$ itself. Informally again, the set ω_1 in the above lemma is defined to be the "smallest uncountable ordinal".

To see what this really means, we need a slightly more formal definition of the ordinal numbers—the informal description above is a bit hard to make precise in formal set-theoretic terms. A more concrete description of the ordinal numbers was introduced by Johann von Neumann, and the idea is to regard each ordinal number as a set whose elements are also sets, namely each ordinal is the set of all ordinals that precede it. In particular, we label the empty set $\varnothing$ as 0, identify the natural number 1 with the set $\{0\} = \{\varnothing\}$, identify 2 with the set $\{0, 1\} = \{\varnothing, \{\varnothing\}\}$, identify

$$3 = \{0, 1, 2\} = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}\}$$

and so forth. Although the notation quickly becomes confusing, one can make sense of von Neumann's general definition:

DEFINITION 18.24. A set S is an ordinal number if and only if S is well ordered with respect to set membership and every element of S is also a subset of S.

If this definition makes your head spin, rest assured that I have the same reaction, but the concept of the ordinal numbers does not rely on anything other than the standard axioms of set theory. With this definition in place, one can define ω_1 as the union of all countable ordinals, which is necessarily uncountable since it would otherwise contain itself.

We now use this to construct a Hausdorff space that is path-connected and locally homeomorphic to $\mathbb{R}$ but is not second countable. This space and various related constructions are sometimes referred to as the **long line**. Let

$$L = \omega_1 \times [0, 1),$$

and define a total order on L such that $(x, s) \leqslant (y, t)$ whenever either $x \leqslant y$ or both $x = y$ and $s \leqslant t$ hold. Writing $x < y$ to mean $x \leqslant y$ and $x \neq y$ for $x, y \in L$, the total order determines a natural topology on L, called the **order topology**, whose base is the collection of all "open" intervals

$$(a, b) := \{x \in L \mid a < x < b\}$$

for arbitrary values $a, b \in L$. The proof of the following statement is an amusing exercise for a rainy day.

PROPOSITION 18.25. *Every point of L has a neighborhood homeomorphic to either $\mathbb{R}$ or (in the case of $(0, 0) \in L$) the half-interval $[0, \infty)$. Moreover, L is Hausdorff and is sequentially compact, but not compact; in particular the set $\{(x, 1/2) \mid x \in \omega_1\} \subset L$ is an uncountable discrete subset of L, implying that L cannot be second countable.* □

I'm guessing you find it especially surprising that this enormous space L is sequentially compact, but that has to do with a peculiar property built into the definition of the set ω_1: every sequence in ω_1 has an upper bound. This is almost immediate from the definition of the ordinal numbers, as for any given sequence $x_n \in \omega_1$, the elements x_n are also (necessarily countable) sets of ordinal numbers, hence their union $\bigcup_n x_n$ is another ordinal number and is countable, meaning it is an element of ω_1, and it clearly bounds the sequence from above.

In dimensions $n \geqslant 2$, there are further constructions of non-second countable but locally Euclidean Hausdorff spaces which do not rely on anything so exotic as the ordinal numbers. An example is the *Prüfer surface*; see the exercise below. But I'm only talking about these things now in order to explain why I will never mention them again.

EXERCISE 18.26. The **Prüfer surface** is an example of a space that would be a connected 2-dimensional manifold if we did not require manifolds to be second countable. It is defined as follows: let $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$, and associate to each $a \in \mathbb{R}$ a copy of the plane $X_a := \mathbb{R}^2$. The Prüfer surface is then

$$\Sigma := \mathbb{H} \amalg \left(\coprod_{a \in \mathbb{R}} X_a \right) \Big/ \sim$$

where the equivalence relation identifies each point $(x, y) \in X_a$ for $y > 0$ with the point $(a + yx, y) \in \mathbb{H}$. Notice that $\mathbb{H}$ and X_a for each $a \in \mathbb{R}$ can be regarded naturally as subspaces of Σ.

(a) Prove that Σ is Hausdorff.
(b) Prove that Σ is path-connected.
(c) Prove that every point in Σ has a neighborhood homeomorphic to $\mathbb{R}^2$.
(d) Prove that a second countable space can never contain an uncountable discrete subset. Then find an uncountable discrete subset of Σ.

19. Surfaces and triangulations

As far as I'm aware, dimension one is the only case in which the problem of classifying *arbitrary* (compact or noncompact) manifolds up to homeomorphism has a reasonable solution. In this lecture we will do the next best thing in dimension two: we will classify all *compact* surfaces. We will focus in particular on closed and connected surfaces. The classification of compact connected surfaces with boundary can easily be derived from this (see Exercise 20.13), and of course compact disconnected surfaces are all just disjoint unions of finitely many connected surfaces, so we lose no generality by restricting to the connected case.

Let us first enumerate the closed connected surfaces that we are already familiar with.

EXAMPLES 19.1. The sphere $S^2 = \Sigma_0$ and torus $\mathbb{T}^2 = \Sigma_1$ are both examples of "oriented surfaces of genus g," which can be defined for any nonnegative integer $g \geqslant 0$ and denoted by Σ_g. In particular, we've seen that for each $g \geqslant 1$, Σ_g is homeomorphic to the g-fold connected sum of copies of $\mathbb{T}^2$, and we have also computed its fundamental group

$$\pi_1(\Sigma_g) \cong \left\{ a_1, b_1, \dots, a_g, b_g \;\middle|\; \prod_{i=1}^{g} [a_i, b_i] = e \right\},$$

whose abelianization is isomorphic to $\mathbb{Z}^{2g}$.

EXAMPLES 19.2. An analogous sequence of surfaces can be defined by taking repeated connected sums of copies of $\mathbb{RP}^2$, e.g. $\mathbb{RP}^2 \# \mathbb{RP}^2$ is homeomorphic to the Klein bottle. By the same trick that we used in Lecture 13 to understand Σ_g, the g-fold connected sum $\#_{i=1}^{g} \mathbb{RP}^2$ is homeomorphic to a space obtained from a polygon with $2g$ edges by identifying them in pairs according to the sequence $a_1, a_1, \dots, a_g, a_g$, thus

$$\pi_1 \left(\#_{i=1}^{g} \mathbb{RP}^2 \right) \cong \left\{ a_1, \dots, a_g \mid a_1^2 \dots a_g^2 = e \right\}.$$

EXERCISE 19.3. For $i = 1, \dots, g - 1$, let $e_i \in \mathbb{Z}^{g-1}$ denote the ith standard basis vector. Show that there is a well-defined homomorphism $G := \{a_1, \dots, a_g \mid a_1^2 \dots a_g^2 = e\} \to \mathbb{Z}^{g-1} \oplus \mathbb{Z}_2$ such that

$$a_i \mapsto \begin{cases} (e_i, 0) & \text{for } i = 1, \dots, g - 1, \\ (-1, \dots, -1, 1) & \text{for } i = g, \end{cases}$$

and that it descends to an isomorphism of the abelianization of G to $\mathbb{Z}^{g-1} \oplus \mathbb{Z}_2$.

Appealing to the standard classification of finitely generated abelian groups, we deduce from the above exercise that all of our examples so far are topologically distinct:

LEMMA 19.4. *No two of the closed surfaces listed in Examples 19.1 and 19.2 are homeomorphic.* □

You might now be wondering whether new examples can be constructed by taking the connected sum of a surface from Example 19.1 with some surface from Example 19.2. The answer is no:

PROPOSITION 19.5. $\mathbb{RP}^2 \# \mathbb{T}^2$ *is homeomorphic to the connected sum of* $\mathbb{RP}^2$ *with the Klein bottle.*[26]

PROOF. Given any surface Σ with two disjoint disks removed, one can construct a new surface by attaching a "handle" of the form $[-1, 1] \times S^1$:

$$\Sigma' := \left(\Sigma \backslash (\mathring{\mathbb{D}}^2 \amalg \mathring{\mathbb{D}}^2)\right) \cup_{S^1 \amalg S^1} \left([-1, 1] \times S^1\right).$$

This operation is essentially the same as the connected sum, except we allow the two disks to be embedded (disjointly) into a single surface Σ rather than two separate surfaces; we sometimes call this a "self-connected sum". As with the connected sum, it depends on a choice of embedding

$$i_1 \amalg i_2 : \mathbb{D}^2 \amalg \mathbb{D}^2 \hookrightarrow \Sigma,$$

but only up to homotopy through embeddings, i.e. modifying the embedding through a continuous 1-parameter family of embeddings will change Σ' into something homeomorphic to the original Σ'.

Let us now shift our perspective on the operation that changes Σ into Σ'. For this it would be helpful to have some pictures, and I do not have time to draw them, but I recommend having a look at Figure 1 in [**FW99**]. Suppose the two holes you're drilling in Σ are right next to each other, but before you drill them, you push the surface up a bit from underneath, creating a disk-shaped lump. Now pick two smaller disk-shaped areas within that lump and push those up even further. Then drill the holes in those two places and attach the handle. We haven't changed any of the topology in creating these "lumps," but we have changed the picture, and if you're imagining it the way that I intended, it now looks like instead of cutting out two holes and attaching a handle, you cut out *one* hole (the base of the original lump) and attached $\Sigma_{1,1}$, the torus with a disk removed. In other words, you performed the connected sum of Σ with $\mathbb{T}^2$:

$$\Sigma' \cong \Sigma \# \mathbb{T}^2.$$

So far so good. . . now let's modify the procedure once more. Viewing $\mathbb{D}^2$ as the unit disk in $\mathbb{C}$, let's replace one of our embeddings $i_1 : \mathbb{D}^2 \to \Sigma$ with another one that has the same image but changes the parametrization by complex conjugation:

$$i_1' : \mathbb{D}^2 \hookrightarrow \Sigma : z \mapsto i_1(\bar{z}).$$

While we will now be cutting out the same two holes in Σ, the way that we attach the handle at the first hole needs to change because $i_1'|_{\partial \mathbb{D}^2}$ parametrizes the circle in the opposite direction from $i_1|_{\partial \mathbb{D}^2}$. The effect is the same as if you were to cut open Σ' along the circle at the boundary of the first hole, flip it's orientation and then glue it back together. Unfortunately you cannot do this in 3-dimensional space—for the same reasons that you cannot embed a Klein bottle into $\mathbb{R}^3$—but it's easy to define the topological space that results from this modification. The effect is precisely to replace the torus in the above description of a connected sum with the Klein bottle; if we call Σ'' the space that results from attaching the handle along this modified gluing map, we have

$$\Sigma'' \cong \Sigma \# K^2,$$

where K^2 denotes the Klein bottle.

---

[26] This proposition has its very own Youtube video, see `https://www.youtube.com/watch?v=aBbDvKq4JqE&t=20s`. Maybe you'll find it helpful. . . I'm not entirely sure if I did.

Finally, let's specify this to the case $\Sigma = \mathbb{RP}^2$. The projective plane has a special property that many surfaces don't: it contains an embedded Möbius band, call it $\mathbb{M}$. Now suppose we construct $\mathbb{RP}^2 \# \mathbb{T}^2$ by embedding two small disks disjointly into $\mathbb{M} \subset \mathbb{RP}^2$, then cutting both out and gluing in a handle. By the previous remarks, the homeomorphism type of the resulting surface will not change if we now move the first hole continuously along a circle traversing $\mathbb{M}$, and the orientation reversal as we traverse $\mathbb{M}$ thus allows us to deform $i_1 : \mathbb{D}^2 \hookrightarrow \mathbb{RP}^2$ to $i_1' : \mathbb{D}^2 \hookrightarrow \mathbb{RP}^2$ through a continuous family of embeddings disjoint from the second disk. This proves that if $\Sigma = \mathbb{RP}^2$, then the two surfaces Σ' and Σ'' described above are homeomorphic. $\square$

It is sometimes useful to make a distinction between two types of handle attachment that were described in the above proof. In one case, the two holes $\mathbb{D}^2 \hookrightarrow \Sigma$ are embedded "right next to each other" and with opposite orientations—in precise terms, this means we focus on the domain of a single chart on Σ, assume both holes are in this domain, define i_1' by translating the image of i_2 in some direction to make it disjoint, and then define $i_1(z) = i_1'(\bar{z})$. The handle attachment that results is straightforward to draw, see e.g. Figure 1 in [**FW99**]. If we then leave the positions of the two holes the same but reverse an orientation by replacing i_1 with i_1', the handle attachment can no longer be embedded in $\mathbb{R}^3$, though this does not stop some authors from trying to draw pictures of it anyway (see Figure 2 in [**FW99**]). This type of handle attachment is sometimes referred to as a **cross-handle**. One should not take this terminology too seriously since the main point of the above prove was that in certain cases such as $\Sigma = \mathbb{RP}^2$, there is no globally meaningful distinction between ordinary handles and cross-handles, i.e. if the two holes do not lie in the same chart, it is not always possible to say that we are dealing with one type of handle and not the other. The distinction does make sense however if both holes are in the same chart, so we will occasionally also use the term "cross-handle" in this situation.

Proposition 19.5 told us that the most obvious way to produce new examples of closed connected surfaces out of the inventory in Examples 19.1 and 19.2 does not actually give anything new. The reason for this turns out to be that there are no others:

THEOREM 19.6. *Every closed connected surface is homeomorphic to either Σ_g for some $g \geqslant 0$ or $\#_{i=1}^g \mathbb{RP}^2$ for some $g \geqslant 1$, where the integer g is in each case unique.*

The uniqueness in this statement already follows from the computations of fundamental groups explained above, so in light of Proposition 19.5, we only still need to show that every closed connected surface other than the sphere is homeomorphic to something constructed out of copies of $\mathbb{T}^2$ and $\mathbb{RP}^2$ by connected sums. (Note that whenever both $\mathbb{T}^2$ and $\mathbb{RP}^2$ appear in this collection, Prop. 19.5 allows us to replace $\mathbb{T}^2$ with two copies of $\mathbb{RP}^2$, as $\mathbb{RP}^2 \# \mathbb{RP}^2$ is the Klein bottle.) We will sketch a proof of this below that is due to John Conway and known colloquially as Conway's "ZIP proof". Another readable account of it is given in [**FW99**].

To frame the problem properly, let us say that for Σ a compact (but not necessarily closed or connected) surface, Σ is *ordinary* if there is a finite sequence of compact surfaces

$$\Sigma^{(0)}, \Sigma^{(1)}, \dots, \Sigma^{(m)} = \Sigma$$

such that $\Sigma^{(0)}$ is a finite disjoint union of spheres $\coprod_{i=1}^N S^2$, and each $\Sigma^{(j+1)}$ is homeomorphic to something obtained from $\Sigma^{(j)}$ by performing one of the following operations:

(1) Removing an open disk from the interior, i.e.

$$\Sigma^{(j+1)} \cong \Sigma^{(j)} \backslash \mathring{\mathbb{D}}^2$$

for some embedding $\mathbb{D}^2 \hookrightarrow \Sigma^{(j)} \backslash \partial \Sigma^{(j)}$;

(2) Attaching a handle (or "cross-handle") to connect two separate boundary components $\ell_1, \ell_2 \subset \partial\Sigma^{(j)}$, i.e.
$$\Sigma^{(j+1)} \cong \Sigma^{(j)} \cup_{\ell_1 \amalg \ell_2} ([-1,1] \times S^1)$$
for some choice of homeomorphism $\partial([-1,1] \times S^1) = S^1 \amalg S^1 \to \ell_1 \amalg \ell^2$;

(3) Attaching a disk (called a **cap**) to a boundary component $\ell \subset \partial\Sigma^{(j)}$, i.e.
$$\Sigma^{(j+1)} \cong \Sigma^{(j)} \cup_\ell \mathbb{D}^2$$
for some choice of homeomorphism $\partial\mathbb{D}^2 = S^1 \to \ell$;

(4) Attaching a Möbius band (called a **cross-cap**) $\mathbb{M}$ to a boundary component $\ell \subset \partial\Sigma^{(j)}$, i.e.
$$\Sigma^{(j+1)} \cong \Sigma^{(j)} \cup_\ell \mathbb{M}$$
for some choice of homeomorphism $\partial\mathbb{M} \cong S^1 \to \ell$.

The classification of 1-manifolds is implicitly in the background of the last three operations: since $\Sigma^{(j)}$ is a compact 2-manifold, $\partial\Sigma^{(j)}$ is a closed 1-manifold and is therefore always a finite disjoint union of circles. Observe now that each of the operations can be reinterpreted in terms of connected sums, e.g. cutting out two holes and then attaching a handle or cross-handle is equivalent to taking the connected sum with $\mathbb{T}^2$ or $\mathbb{RP}^2\#\mathbb{RP}^2$, while attaching a cap or cross-cap gives connected sums with S^2 or $\mathbb{RP}^2$ respectively. It follows that any ordinary surface that is also closed and connected necessarily belongs to our existing inventory of closed and connected surfaces, thus it will suffice to prove:

LEMMA 19.7. *Every closed surface is ordinary.*

At this point in almost every topology class, it becomes necessary to cheat a bit and appeal to a fundamental result about surfaces that is believable and yet far harder to prove than we have time to discuss in any detail. I'm referring to the existence of *triangulations*. This is not only a useful tool in classifying surfaces, but also will play a large motivational role when we introduce homology. The following is thus simultaneously a necessary digression behind the proof of Lemma 19.7 and also a preview of things to come.

The idea of a triangulation is to decompose a topological n-manifold into many homeomorphic pieces that we think of as "n-dimensional triangles". More precisely, the **standard n-simplex** is defined as the set
$$\Delta^n := \left\{(t_0, \dots, t_n) \in I^{n+1} \;\middle|\; t_0 + \ldots + t_n = 1\right\}$$
for each integer $n \geqslant 0$. This makes Δ^0 the one-point space $\{1\} \subset \mathbb{R}$, while Δ^1 is a compact line segment in $\mathbb{R}^2$ homeomorphic to the interval I, Δ^2 is the compact region in a plane bounded by a triangle, Δ^3 is the compact region in a 3-dimensional vector space bounded by a tetrahedron, and so forth. For a surface Σ, we would now like to view copies of Δ^2 as fundamental building blocks of Σ, arranged in such a way that the intersection between any two of those building blocks is either empty or is a copy of Δ^1 or Δ^0. One can express this condition in purely combinatorial terms by thinking of Δ^n as the convex hull of its $n+1$ *vertices*, which are the standard basis vectors of $\mathbb{R}^{n+1}$. In this way, an n-simplex is always determined by $n+1$ vertices, and this idea can be formalized via the notion of a *simplicial complex*.

DEFINITION 19.8. A **simplicial complex** (*Simplizialkomplex*) K consists of two sets V and S, called the sets of **vertices** (*Eckpunkte*) and **simplices** (*Simplizes*) respectively, where the elements of S are nonempty finite subsets of V, and $\sigma \in S$ is called an **n-simplex** of K if it has $n+1$ elements. We require the following conditions:

(1) Every vertex $v \in V$ gives rise to a 0-simplex in K, i.e. $\{v\} \in S$;
(2) If $\sigma \in S$ then every subset $\sigma' \subset \sigma$ is also an element of S.

For any n-simplex $\sigma \in S$, its subsets are called its **faces** (*Seiten* or *Facetten*), and in particular the subsets that are $(n-1)$-simplices are called **boundary faces** (*Seitenflächen*) of σ. The second condition above thus says that for every simplex in the complex, all of its boundary faces also belong to the complex. With this condition in place, the first condition is then equivalent to the requirement that every vertex in the set V belongs to at least one simplex.

The complex K is said to be **finite** if V is finite, and it is n**-dimensional** if

$$\sup_{\sigma \in S} |\sigma| = n + 1,$$

i.e. n is the largest number for which K contains an n-simplex.

Though the definition above is purely combinatorial, there is a natural way to associate a topological space $|K|$ to any simplicial complex K. We shall describe it only in the case of a finite complex,[27] since that is what we need for our discussion of compact surfaces. Given $K = (V, S)$, choose a numbering of the vertices $V = \{v_1, \ldots, v_N\}$ and associate to each k-simplex $\sigma = \{v_{i_0}, \ldots, v_{i_k}\}$ the set

$$\Delta_\sigma := \left\{ (t_1, \ldots, t_N) \in I^N \;\middle|\; t_{i_0} + \ldots + t_{i_k} = 1 \text{ and } t_j = 0 \text{ for all } v_j \notin \sigma \right\}.$$

Notice that Δ_σ is homeomorphic to the standard k-simplex Δ^k, but lives in the subspace of $\mathbb{R}^N$ spanned by the specific coordinates corresponding to its vertices. The **polyhedron** (*Polyeder*) of K is then the compact space

$$|K| := \bigcup_{\sigma \in S} \Delta_\sigma \subset \mathbb{R}^N.$$

While the definition above makes $|K|$ a subset of a Euclidean space that may have very large dimension in general, it is not so hard to picture $|K|$ in a few simple examples.

EXAMPLE 19.9. Suppose $V = \{v_0, v_1, v_2\}$ and S is defined to consist of all subsets of V. Then $|K|$ is just the standard 2-simplex Δ^2.

EXAMPLE 19.10. Suppose $V = \{v_0, v_1, v_2, v_3\}$ and S contains the subsets $A := \{v_0, v_1, v_2\}$ and $B := \{v_1, v_2, v_3\}$, plus all of their respective subsets. Then $|K|$ contains two copies of the triangle Δ^2, which we can label A and B, and they intersect each other along a single common edge connecting the vertices labeled v_1 and v_2. In particular, $|K|$ is homeomorphic to a 2-dimensional square I^2, formed by gluing two triangles together along one edge.

DEFINITION 19.11. A **triangulation** (*Triangulierung*) of a compact topological n-manifold M is a homeomorphism of M to the polyhedron of a finite n-dimensional simplicial complex.

In particular, this makes precise the notion of decomposing a surface Σ into triangles (copies of Δ^2) whose intersections with each other are always simplices of lower dimension. Observe that in a triangulated surface Σ with $\partial\Sigma = \emptyset$, the fact that every point in one of the 1-simplices σ has a neighborhood homeomorphic to $\mathbb{R}^2$ implies that σ is a boundary face of *exactly two* 2-simplices in the triangulation. One can say the same about the $(n-1)$-simplices in any triangulation of a closed n-manifold. This is not a property that arbitrary simplicial complexes have, but it is a general property of the complexes that appear in triangulations of closed manifolds.

THEOREM 19.12. *Every closed surface admits a triangulation.*

---

[27] The polyhedron of a finite simplicial complex has an obvious topology because it comes with an embedding into some finite-dimensional Euclidean space. For infinite complexes this is not true, and thus more thought is required to define the *right* topology on $|K|$. We would need to talk about this if we wanted to define triangulations of noncompact spaces, but since we don't want that right now, we will not. The correct topology on infinite complexes will be discussed next semester; see 29.

This theorem is old enough for the first proof to have been published in German [**Rad25**], and it was not the main result of the paper in which it appeared, yet it is in some sense far harder than it has any right to be—it seems to be one of the rare instances in mathematics where learning cleverer high-powered techniques does not really help. I can at least sketch what is involved. Since a closed surface Σ can be covered by finitely many charts, it can also be covered by a finite collection of regions homeomorphic to $\mathbb{D}^2$, which is homeomorphic to the standard 2-simplex Δ^2. Of course the interiors of these 2-simplices overlap, which is not allowed in a triangulation, but the idea is to examine each of the overlap regions and subdivide it further into simplices. By "overlap region," what I mean is the following: if $D_1, \dots, D_N \subset \Sigma$ denote the finite collection of disks $D_i \cong \Delta^2$ covering Σ, whose boundaries are loops ∂D_i, then the closure of each connected component of $\Sigma \backslash \bigcup_i \partial D_i$ is a region that needs to be subdivided into triangles. After perturbing each of the disks D_i so that its boundary intersects the other boundaries only finitely many times, we can arrange for each of these overlap regions to be bounded by embedded circles, and notice that since each of the regions is contained in at least one of the disks D_i, we can view them as subsets of $\mathbb{R}^2$. Now, I don't know about you, but I find it not so hard to believe that regions in $\mathbb{R}^2$ bounded by embedded circles can be subdivided into triangles in a reasonable way—I would imagine that writing down a complete algorithm to do this is a pain in the neck, but it sounds plausible. It may surprise you however to know that it is very far from obvious what the region bounded by an embedded circle in $\mathbb{R}^2$ can look like in general. Actually the answer is simple and is what you would expect: the region is homeomorphic to a disk, but this is not at all easy to prove, it is an important theorem in classical topology known as the *Schönflies theorem*. With this result in hand, one can formulate an algorithm for triangulating surfaces as sketched above by triangulating the disk-like overlap regions. Complete accounts of this are given in [**Moi77**] and [**Tho92**].

Note that if Σ is not just a topological 2-manifold but also has a *smooth* structure, then one can avoid the Schönflies theorem by appealing to some basic facts from Riemannian geometry. Choosing a Riemannian metric allows us to define the notion of a "straight line" (geodesic) on the manifold, and one can arrange in this case for the disks D_i to be convex, so that the overlap regions are also convex and therefore obviously homeomorphic to disks. This trick actually works in arbitrary dimensions, leading to the result that *smooth* manifolds can be triangulated in any dimension. For topological manifolds this is not true in general: it is true in dimension three (see [**Moi77**]), but from dimension four upwards there are examples of topological manifolds that do not admit triangulations. The case of dimension five has only been understood since 2013—see [**Man14**] for a readable survey of this subject and its history.

But enough about triangulations: let's just assume that surfaces can be triangulated and use this to finish the classification theorem.

Proof of Lemma 19.7. Assume Σ is a closed surface homeomorphic to the polyhedron $|K|$ of a finite 2-dimensional simplicial complex $K = (V, S)$ with 2-simplices $\sigma_1, \dots, \sigma_N$. By abuse of notation, we shall also denote by $\sigma_1, \dots, \sigma_N$ the corresponding subsets of Σ homeomorphic to the standard 2-simplex Δ^2. The latter is homeomorphic to $\mathbb{D}^2 \cong S^2 \backslash \mathring{\mathbb{D}}^2$, thus

$$\Sigma^{(0)} := \sigma_1 \amalg \dots \amalg \sigma_N$$

is ordinary. The idea now is to reconstruct Σ from this disjoint union by gluing pairs of 2-simplices together along corresponding boundary faces one at a time, producing a sequence of compact surfaces $\Sigma^{(j)}$, each of which may be disconnected and have nonempty boundary except for the last in the sequence, which is Σ. The operation changing $\Sigma^{(j)}$ to $\Sigma^{(j+1)}$ is performed by gluing together two arcs $\ell_1, \ell_2 \subset \partial \Sigma^{(j)}$, i.e. we can write

$$\Sigma^{(j+1)} = \Sigma^{(j)} / \sim \qquad \text{where} \qquad \sim \text{ identifies } \ell_1 \text{ with } \ell_2,$$

with ℓ_1 and ℓ_2 assumed to be individual boundary faces of two distinct 2-simplices. These boundary faces are each homeomorphic to the compact interval I, and their interiors are disjoint subsets of $\Sigma^{(j)}$, but they may have boundary points (vertices of the triangulation) in common if some neighboring pair of corresponding boundary faces has already been glued together in the process of turning $\Sigma^{(0)}$ into $\Sigma^{(j)}$. One can now imagine various scenarios, based on the knowledge (thanks to the classification of 1-manifolds) that every connected component of $\partial\Sigma^{(j)}$ is a circle:

Case 1: $\ell_1 \cup \ell_2$ forms a single connected component of $\partial\Sigma^{(j)}$. Gluing them together is then equivalent to attaching either a cap or a cross-cap to that boundary component, depending on the orientation of the homeomorphism that identifies them.

Case 2: ℓ_1 and ℓ_2 form part of a single connected component of $\partial\Sigma^{(j)}$, but not all of it, i.e. their boundary vertices are not exactly the same, so that there are either one or two gaps between them forming additional arcs on some circle in $\partial\Sigma^{(j)}$. Gluing them together then is equivalent to attaching a cap or cross-cap as in case 1, except that it leaves one or two holes where the gaps were, so we can realize this operation by attaching the cap/cross-cap and drilling holes afterward.

Case 3: ℓ_1 and ℓ_2 lie on different connected components of $\partial\Sigma^{(j)}$. Then neither can be the entirety of a boundary component since both are homeomorphic to I instead of S^1, though it's useful to imagine what would happen if both really were the entirety of a boundary component: gluing them together would then be equivalent to attaching a handle. The useful way to turn this picture into reality is to imagine both ℓ_1 and ℓ_2 as making up *most* of their respective boundary components, each leaving a very small gap where their end points fail to come together. Gluing ℓ_1 to ℓ_2 is then equivalent to attaching a handle but then drilling a small hole in it.

In all of these cases, the operation that converts $\Sigma^{(j)}$ into $\Sigma^{(j+1)}$ can be realized by a finite sequence of operations from our stated list, so carrying out this procedure as many times as necessary to convert $\Sigma^{(0)}$ into Σ produces a surface that is ordinary. $\square$

Exercise 19.13. Recall that if Σ is a surface with boundary, the **boundary** $\partial\Sigma$ is defined as the set of all points $p \in \Sigma$ such that some chart $\varphi : \mathcal{U} \xrightarrow{\cong} \Omega \subset \mathbb{H}^2$ defined on a neighborhood $\mathcal{U} \subset \Sigma$ of p satisfies $\varphi(p) \in \partial\mathbb{H}^2$. Here $\mathbb{H}^2 := [0,\infty) \times \mathbb{R} \subset \mathbb{R}^2$, $\partial\mathbb{H}^2 := \{0\} \times \mathbb{R} \subset \mathbb{H}^2$, and Ω is an open subset of $\mathbb{H}^2$. One can analogously define $p \in \Sigma$ to be an *interior point* of Σ of some chart maps it to $\mathbb{H}^2 \backslash \partial\mathbb{H}^2$. Prove that no point on $\partial\Sigma$ is also an interior point of Σ.
Hint: If you have two charts defined near p such that one sends p to $\partial\mathbb{H}^2$ while the other sends it to $\mathbb{H}^2 \backslash \partial\mathbb{H}^2$, then a transition map relating these two charts maps some neighborhood in $\mathbb{H}^2$ of a point $x \in \mathbb{H}^2 \backslash \partial\mathbb{H}^2$ to a neighborhood in $\mathbb{H}^2$ of a point $y \in \partial\mathbb{H}^2$. What happens to this homeomorphism if you remove the points x and y? Think about the fundamental group.
Remark: A similar result is true for topological manifolds of arbitrary dimension, but you do not yet have enough tools at your disposal to prove this. A proof using singular homology will be possible before the end of the semester.

Exercise 19.14. This exercise concerns manifolds with smooth structures, which were discussed briefly in Lecture 18 (see especially Definition 18.10 and Theorem 18.11). We will need the following additional notions:

- For two smooth manifolds M and N, a map $f : M \to N$ is called **smooth** if for every pair of smooth charts ψ_β on N and φ_α on M, the map $f_{\beta\alpha} := \psi_\beta \circ f \circ \varphi_\alpha^{-1}$ is C^∞ wherever it is defined. (In other words, f is "C^∞ in local coordinates".)
- For $f : M \to N$ a smooth map between smooth manifolds, a point $q \in N$ is a **regular value** of f if for all charts φ_α on M and ψ_β on N such that q is in the domain of ψ_β, $\psi_\beta(q)$ is a regular value of $f_{\beta\alpha}$. (In other words, q is a "regular value of f in local coordinates".)

An easy corollary of the usual implicit function theorem (Theorem 18.11) then states that if M is a smooth m-manifold without boundary, N is a smooth n-manifold and $f : M \to N$ is a smooth map that has $q \in N$ as a regular value, the preimage $f^{-1}(q) \subset M$ is a smooth submanifold[28] of dimension $m-n$. If M has boundary, then one should assume additionally that q is a regular value of the restricted map $f|_{\partial M} : \partial M \to N$, and the conclusion is then that $Q := f^{-1}(q)$ is a smooth manifold of dimension $m - n$ with boundary $\partial Q = Q \cap \partial M$.

We will use the following perturbation lemma as a block box: if M and N are compact smooth manifolds, $q \in N$ and $f : M \to N$ is continuous, then every neighborhood of f in $C(M, N)$ with the compact-open topology (cf. Exercise 7.28) contains a smooth map $f_\epsilon : M \to N$ for which q is a regular value of both f_ϵ and $f_\epsilon|_{\partial M}$. Moreover, if $f|_{\partial M}$ is already smooth and has q as a regular value, then the perturbation can be chosen such that $f_\epsilon|_{\partial M} = f|_{\partial M}$. Proofs of these statements can be found in standard books on differential topology such as [**Hir94**].

If you take all of this as given, then you can use it to define something quite beautiful. Assume M and N are closed connected smooth manifolds of the same dimension n. Then for any smooth map $f : M \to N$ with regular value $q \in N$, the implicit function theorem implies that $f^{-1}(q)$ is a compact 0-manifold, i.e. a finite set of points. Define the **mod 2 mapping degree** $\deg_2(f) \in \mathbb{Z}_2$ of f by

$$\deg_2(f) := |f^{-1}(q)| \pmod 2,$$

i.e. $\deg_2(f)$ is $0 \in \mathbb{Z}_2$ if the number of points in $f^{-1}(q)$ is even, and $1 \in \mathbb{Z}_2$ if it is odd.

(a) Prove that for any given choice of the point $q \in N$, the degree $\deg_2(f) \in \mathbb{Z}_2$ depends only on the homotopy class of the map $f : M \to N$.
Hint: If you have a homotopy $H : I \times M \to N$ between two maps, perturb it as necessary and look at $H^{-1}(q)$. Use the classification of compact 1-manifolds.
Remark: One can show with a little more effort that $\deg_2(f)$ also does not depend on the choice of the point q, and moreover, it has a well-defined extension to continuous (but not necessarily smooth) maps $f : M \to N$, defined by setting $\deg_2(f) := \deg_2(f_\epsilon)$ for any sufficiently close smooth perturbation f_ϵ that has q as a regular value.

(b) Prove that every continuous map $f : S^2 \to S^2$ homotopic to the identity is surjective.

(c) What goes wrong with this discussion of we allow M to be a noncompact manifold? Describe two homotopic maps $f, g : \mathbb{R} \to S^1$ for which $\deg_2(f)$ and $\deg_2(g)$ can be defined in the manner described above but are not equal.

(d) Prove that if $n > m$, every continuous map $S^m \to S^n$ is homotopic to a constant map.
Hint: What does it mean for a point $q \in S^n$ to be a regular value of $f : S^m \to S^n$ if $n > m$?

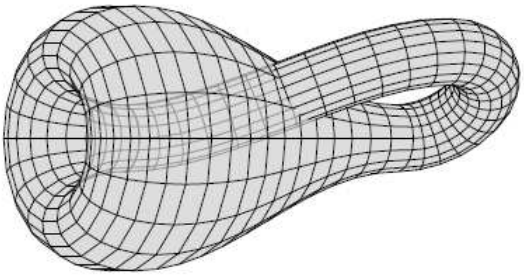
20. Orientations

This lecture is in part an addendum to the classification of surfaces, though it will also introduce some concepts that will be useful to have in mind when we discuss homology.

I have used the word “orientation” many times in this course without giving any precise explanation of what it means. I want to do that now, at least for manifolds of dimensions one and two. The canonical example to have in mind is the Klein bottle:

---

[28] A subset $Y \subset M$ of a smooth m-manifold M is called a **smooth submanifold** (*glatte Untermannigfaltigkeit*) of dimension k if every point $p \in Y$ has a neighborhood $\mathcal{U} \subset M$ admitting a so-called **slice chart** (*Bügelkarte*), meaning a smooth chart $\varphi : \mathcal{U} \to \mathbb{R}^n$ with the property that $Y \cap \mathcal{U} = \varphi^{-1}(\mathbb{R}^k \times \{0\})$. Covering Y with slice charts then gives Y the structure of a smooth k-manifold for which the inclusion $Y \hookrightarrow M$ is a smooth map. As an important special case: the boundary $\partial M \subset M$ of a smooth m-manifold is always a smooth $(m-1)$-dimensional submanifold.



This standard picture of the Klein bottle is unfortunately the image of a *non-injective* map $i : K^2 \to \mathbb{R}^3$ into 3-dimensional Euclidean space from a certain closed 2-manifold K^2: in differential geometry, one would call $i : K^2 \to \mathbb{R}^3$ an *immersion*, which fails to be an *embedding* (and its image is therefore not a *submanifold* of $\mathbb{R}^3$) because one can see a pair of disjoint circles $C_1, C_2 \subset K^2$ such that $i(C_1) = i(C_2)$. For the following informal discussion, however, let us ignore this detail and pretend that $i : K^2 \to \mathbb{R}^3$ is an embedding, with no self-intersections.[29] Now, aside from the fact that it cannot be embedded into $\mathbb{R}^3$, what most of us really find strange about the Klein bottle is that we cannot make a meaningful distinction between the "inside" and the "outside" of the surface. If, for instance, you were an insect and somebody tried to trap you inside a glass Klein bottle, then you could just walk along the surface until you are standing on the opposite side of the glass, and you are free. In mathematical terms, this means that the Klein bottle $K^2 \subset \mathbb{R}^3$ admits an embedded loop $\gamma : I \to K^2$ along which a continuous family of nonzero vectors $V(t) \in \mathbb{R}^3$ can be found which are orthogonal to the surface at each $\gamma(t)$ and satisfy $V(1) = -V(0)$. By contrast, if you take any embedded loop $\gamma : I \to \mathbb{T}^2 \subset \mathbb{R}^3$ on the torus in its standard representation as a tube-like subset of $\mathbb{R}^3$, and choose a normal vector field $V(t)$ along this loop, $V(1)$ will always need to be a positive multiple of $V(0)$. That's because there *is* a meaningful distinction between the outside and inside of the torus $\mathbb{T}^2 \subset \mathbb{R}^3$.[30]

But this discussion of "inside" vs. "outside" is not really satisfactory, because whenever we talk about normal vectors, we are referring to a piece of data that is not intrinsic to the spaces $\mathbb{T}^2$ or K^2. It depends rather on how we choose to embed or immerse them in $\mathbb{R}^3$. So how can we talk about orientations without mentioning normal vectors?

To answer this, imagine again that you are an insect standing on the surface of the Klein bottle, and while standing in place, you turn around in a circle, rotating 360 degrees to your left. An observer from the outside will see you turn, but the *direction* of the turn that observer sees will depend on which side of the glass you are standing on. In particular, if you turn around like this and then follow the aforementioned path to come back to the same point but on the other side of the glass, then when you turn again 360 degrees to the left, the outside observer will see you turning the other way. We can use this turning idea to formulate a precise notion of orientation without mentioning normal vectors.

Informally, let us agree that an orientation of a surface should mean a choice of which kinds of rotations at each point are to be labeled "clockwise" as opposed "counterclockwise". This is still not a precise mathematical definition, but now we are making progress. The term "counterclockwise rotation" has a precise and canonical definition in $\mathbb{R}^2$, for instance, thus we can agree that $\mathbb{R}^2$ has a canonical orientation. The natural thing to do is then to use charts to define orientations

---

[29] Notice that if we were willing to map K^2 into $\mathbb{R}^4$ instead of $\mathbb{R}^3$, then we could easily turn i into an injective map $K^2 \hookrightarrow \mathbb{R}^4$ just by slightly perturbing the fourth coordinate along C_1 but not along C_2.

[30] The fancy way of saying this in differential-geometric language is that the *normal bundle* of the standard immersion $K^2 \looparrowright \mathbb{R}^3$ is nontrivial, whereas the standard embedding $\mathbb{T}^2 \hookrightarrow \mathbb{R}^3$ has trivial normal bundle. If you don't know what that means, don't worry about it for now.

on a surface Σ via their local identifications with $\mathbb{R}^2$. There's just one obvious problem with this idea: if all charts are allowed, then the definition of an orientation at some point might depend on our choice of chart to use near that point, because the transition map relating two charts might interchange counterclockwise and clockwise rotations. It therefore becomes important to restrict the class of allowed charts so that transition maps do not change orientations, i.e. so that they are *orientation preserving*. Our main task is to give the latter term a precise definition, and this can be done in terms of winding numbers.

Recall the following notion from Exercise 10.27. For $z \in \mathbb{C}$ and $\epsilon > 0$, define a counterclockwise loop about z by

$$\gamma_{z,\epsilon} : S^1 \hookrightarrow \mathbb{C} : e^{i\theta} \mapsto z + \epsilon e^{i\theta}.$$

Note that for fixed $z \in \mathbb{C}$, varying the value of $\epsilon > 0$ does not change the homotopy class of this loop in $\mathbb{C}\backslash\{z\}$, and for a suitable choice of base point it is always a generator of $\pi_1(\mathbb{C}\backslash\{z\}) \cong \mathbb{Z}$. For $k \in \mathbb{Z}$, define also the loop

$$\gamma^k_{z,\epsilon} : S^1 \to \mathbb{C} : e^{i\theta} \mapsto z + \epsilon e^{ki\theta},$$

which covers $\gamma_{z,\epsilon}$ exactly k times if $k > 0$, covers it $|k|$ times with reversed orientation if $k < 0$, and is constant if $k = 0$. Now for any other loop $\alpha : S^1 \to \mathbb{C}\backslash\{z\}$, the **winding number** (*Windungszahl*) of α about z is an integer characterized uniquely by the condition

$$\operatorname{wind}(\alpha; z) = k \qquad \Longleftrightarrow \qquad \alpha \underset{h}{\sim} \gamma^k_{z,\epsilon} \quad \text{in} \quad \mathbb{C}\backslash\{z\}.$$

If $\mathcal{U}, \mathcal{V} \subset \mathbb{C}$ are open subsets and $f : \mathcal{U} \to \mathcal{V}$ is a homeomorphism, then for any $z \in \mathcal{U}$ with $f(z) = w \in \mathcal{V}$, we can assume the loop $\gamma_{z,\epsilon}$ lies in $\mathcal{U}$ for all $\epsilon > 0$ sufficiently small, and the fact that f is bijective makes $f \circ \gamma_{z,\epsilon}$ a loop in $\mathbb{C}\backslash\{w\}$. It follows that there is a well-defined winding number $\operatorname{wind}(f \circ \gamma_{z,\epsilon}; w) \in \mathbb{Z}$, and shrinking $\epsilon > 0$ to a smaller number $\epsilon' > 0$ obviously will not change it since $\gamma_{z,\epsilon}$ and $\gamma_{z,\epsilon'}$ are homotopic in $\mathcal{U}\backslash\{z\}$, so that $f \circ \gamma_{z,\epsilon}$ and $f \circ \gamma_{z,\epsilon'}$ are homotopic in $\mathbb{C}\backslash\{w\}$.

LEMMA 20.1. *In the situation described above,* $\operatorname{wind}(f \circ \gamma_{z,\epsilon}; w)$ *is always either* 1 *or* -1.

PROOF. Choose $\epsilon > 0$ small enough so that the image of $f \circ \gamma_{z,\epsilon}$ lies in a ball $B_r(w)$ about w with radius $r > 0$ sufficiently small such that $B_r(w) \subset \mathcal{V}$. Then for $\delta \in (0, r)$, the homotopy class of $\gamma_{w,\delta}$ generates $\pi_1(B_r(w)\backslash\{w\}) \cong \pi_1(\mathbb{C}\backslash\{w\}) \cong \mathbb{Z}$, and $k := \operatorname{wind}(f \circ \gamma_{z,\epsilon}; w)$ is the unique integer such that $f \circ \gamma_{z,\epsilon}$ is homotopic in $B_r(w)\backslash\{w\}$ to $\gamma^k_{w,\delta}$. Since $\gamma_{z,\epsilon}$ generates $\pi_1(\mathbb{C}\backslash\{z\})$, there is also a unique integer $\ell \in \mathbb{Z}$ such that $f^{-1} \circ \gamma_{w,\delta}$ is homotopic in $\mathbb{C}\backslash\{z\}$ to $\gamma^\ell_{z,\epsilon}$. This implies

$$\gamma_{z,\epsilon} = f^{-1} \circ f \circ \gamma_{z,\epsilon} \underset{h}{\sim} f^{-1} \circ \gamma^k_{w,\delta} \underset{h}{\sim} \gamma^{k\ell}_{z,\epsilon} \quad \text{in} \quad \mathbb{C}\backslash\{z\},$$

hence $k\ell = 1$. Since k and ℓ are both integers, we conclude both are ± 1. □

EXERCISE 20.2. Show that in the setting of Lemma 20.1, the subsets $\mathcal{U}_\pm = \{z \in \mathcal{U} \mid \operatorname{wind}(f \circ \gamma_{z,\epsilon}; f(z)) = \pm 1\}$ are each both open and closed, so in particular, the sign of this winding number is constant on each connected component of $\mathcal{U}$.
Hint: Since the two sets are complementary, it suffices to prove both are open. What happens to $\operatorname{wind}(f \circ \gamma_{z,\epsilon}; w)$ *if you perturb* z *and* w *independently of each other by very small amounts?*

One can define winding numbers just as well for loops in $\mathbb{R}^2$ by identifying $\mathbb{R}^2$ with $\mathbb{C}$ via $(x, y) \leftrightarrow x + iy$. We have been using complex numbers purely for notational convenience, but in the following we will refer instead to domains in $\mathbb{R}^2$ or the half-plane $\mathbb{H}^2$. The discussion also makes sense for homeomorphisms between open subsets of $\mathbb{H}^2$ as long as we only consider points z in the interior $\mathbb{H}^2\backslash\partial\mathbb{H}^2$, since the loop $\gamma_{z,\epsilon}$ is then contained in $\mathbb{H}^2$ for ϵ sufficiently small. Note that by Exercise 19.13, a homeomorphism between open subsets of $\mathbb{H}^2$ always maps points in $\partial\mathbb{H}^2$ to $\partial\mathbb{H}^2$ and points in $\mathbb{H}^2\backslash\partial\mathbb{H}^2$ to $\mathbb{H}^2\backslash\partial\mathbb{H}^2$.

Definition 20.3. Given open subsets $\mathcal{U}, \mathcal{V} \subset \mathbb{H}^2$, a homeomorphism $f : \mathcal{U} \to \mathcal{V}$ is called **orientation preserving** (*orientierungserhaltend*) if $\operatorname{wind}(f \circ \gamma_{z,\epsilon}; f(z)) = 1$ for all $z \in \mathbb{H}^2 \backslash \partial \mathbb{H}^2$ and $\epsilon > 0$ sufficiently small. It is called **orientation reversing** (*orientierungsumkehrend*) if $\operatorname{wind}(f \circ \gamma_{z,\epsilon}; f(z)) = -1$ for all $z \in \mathbb{H}^2 \backslash \partial \mathbb{H}^2$ and $\epsilon > 0$ sufficiently small.

Lemma 20.1 and Exercise 20.2 together imply that a homeomorphism is always either orientation preserving or orientation reversing on each individual connected component. Similar notions can also be defined in all positive dimensions, not only dimension two, though one needs to replace winding numbers with a different way of measuring the local behavior of a homeomorphism in higher dimensions. In dimension one, the proper definition is fairly obvious:

Definition 20.4. Given open subsets $\mathcal{U}, \mathcal{V}$ in $\mathbb{R}$ or $\mathbb{H} := [0, \infty)$, a homeomorphism $f : \mathcal{U} \to \mathcal{V}$ is called **orientation preserving** if it is an increasing function, and **orientation reversing** if it is a decreasing function.

I will refrain for now from stating the definition for dimensions $n \geqslant 3$, since it requires a certain amount of language (involving degrees of maps between spheres) that we have not yet adequately defined. A more straightforward definition is available however if you are willing to restrict from homeomorphisms to *diffeomorphisms*, i.e. bijections that are C^∞ and have C^∞ inverses. Actually, C^1 is good enough: the point is that the derivative $df(x) : \mathbb{R}^n \to \mathbb{R}^n$ of such a map at any point x is guaranteed to be an *invertible* linear map, so it has a nonzero determinant. One then calls the map orientation preserving if the determinant of its derivative is everywhere positive, and orientation reversing if that determinant is everywhere negative. We will not worry about this in the following since we will almost exclusively talk about orientations for manifolds of dimension at most two. Nonetheless, there is no harm in stating a definition of orientation that is valid for topological manifolds of arbitrary dimension, and the definition will look slightly familiar if you recall our discussion of smooth structures in Lecture 18.

Definition 20.5. An **orientation** (*Orientierung*) of an n-manifold M for $n \geqslant 1$ is a maximal collection of charts $\{\varphi_\alpha : \mathcal{U}_\alpha \to \Omega_\alpha\}_{\alpha \in J}$ such that $M = \bigcup_{\alpha \in J} \mathcal{U}_\alpha$ and all transition maps $\varphi_\beta \circ \varphi_\alpha^{-1}$ are orientation preserving. If M is a 0-manifold, we define an orientation on M to be a function $\epsilon : M \to \{1, -1\}$, which partitions M into sets of positively/negatively oriented points $M_\pm := \epsilon^{-1}(\pm 1)$.

We say that M is **orientable** (*orientierbar*) if it admits an orientation, and refer to any manifold endowed with the extra structure of an orientation as an **oriented manifold** (*orientierte Mannigfaltigkeit*).

Specializing again to dimension 2, an orientation of M allows you to draw small loops around arbitrary points in M and label them "counterclockwise" or "clockwise" in a consistent way, where consistency means in effect that you can never deform a counterclockwise loop continuously through small loops around other points and end up with a clockwise loop. The actual definition of counterclockwise comes from the special collection of charts that an orientation provides: we call these **oriented charts**, and define a small loop about a point in M to be counterclockwise if and only if it looks counterclockwise in an oriented chart.

If M is a 1-manifold, then instead of talking about loops or rotations, we can simply label orientations with arrows: the orientation defines which paths in M can be called "increasing" as opposed to "decreasing".

Remark 20.6. One can show that any orientation-preserving homeomorphism between open subsets of $\mathbb{H}^2$ restricts to the boundary as an orientation-preserving homeomorphism between open subsets of $\partial \mathbb{H}^2 \cong \mathbb{R}$. It follows that there is a natural notion of induced **boundary orientation**, i.e. on any orientable surface Σ with boundary, a choice of orientation on Σ induces a natural orientation on $\partial\Sigma$ by taking the oriented charts on the latter to be restrictions of the oriented charts on Σ. An analogous statement is true for manifolds with boundary in all dimensions. For $\dim M = 1$, one defines the boundary orientation of ∂M by setting $\epsilon(p) = 1$ whenever the "increasing" direction of M points from the interior of M toward the boundary point $p \in \partial M$, and $\epsilon(p) = -1$ whenever this direction points from $p \in \partial M$ toward the interior. (Different authors may define this in slightly different ways, but it usually doesn't matter: the point is just to choose a convention and be consistent about it.)

Let us specialize this discussion to manifolds with triangulations, i.e. manifolds that are homeomorphic to the polyhedron of a simplicial complex. The latter is an essentially combinatorial notion, so orientations of such objects can also be defined in combinatorial terms. Recall that if J is any finite set, any bijection $\pi : J \to J$ is a permutation of its elements, that is, one can identify π with some element of the symmetric S_N group on N objects after choosing a numbering $v_1, \ldots, v_N$ for the elements in J. The symmetric group S_N is generated by ***flips***, meaning permutations that interchange two elements of J while leaving the rest fixed, and we say that $\pi \in S_N$ is an **even** permutation if it can be written as a composition of evenly many flips; otherwise it is an **odd** permutation. If we represent π by an N-by-N matrix permuting the N standard basis vectors of $\mathbb{R}^N$, then we can recognize the even/odd permutations as those for which this matrix has positive/negative determinant respectively; in fact, the matrices of even permutations always have determinant $+1$, and those of odd permutations have determinant -1. To motivate the next definition, recall the definition of the standard n-simplex $\Delta^n = \{(t_0, \ldots, t_n) \mid t_0 + \ldots + t_n = 1\}$. Any element of the symmetric group on $n + 1$ objects can be regarded as a permutation of the vertices of Δ^n numbered from 0 to n, and the matrix representation of this permutation then defines a linear map on $\mathbb{R}^{n+1}$ that permutes the standard basis vectors accordingly. That linear map preserves the subset $\Delta^n \subset \mathbb{R}^{n+1}$, and it is an orientation-preserving transformation on $\mathbb{R}^{n+1}$ if and only if its determinant is positive, which is equivalent to requiring the permutation to be even.

DEFINITION 20.7. For a simplicial complex $K = (V, S)$, an **orientation** of an n-simplex $\sigma \in S$ for $n \geqslant 1$ is an equivalence class of orderings of the vertices $v \in \sigma$, where two orderings are defined to be equivalent if and only if they are related to each other by an even permutation. An orientation of a 0-simplex is defined simply as an assignment of the number $+1$ or -1 to that vertex.

For simplices of dimension 1 or 2 there are easy ways to illustrate in pictures what this definition means; see Figure 11. The figure shows the six possible ways of ordering the three vertices of a 2-simplex, where the individual choices in each row are related to each other by even permutations and thus define equivalent orientations, whereas each choice is related to the one directly underneath it by a single flip, which is an odd permutation. We can represent the orientation itself by drawing a circular arrow that follows the direction of the sequence of vertices labeled $0, 1, 2$, and this arrow depends *only* on the orientation since even permutations of three objects are also *cyclical* permutations.

Another intuitive fact you can infer from Figure 11 is that an orientation of a 2-simplex induces a natural **boundary orientation** for each of its 1-dimensional boundary faces. The latter orientations are represented in the picture by arrows pointing from one vertex to another, meant to indicate the ordering of the two vertices, and the visual recipe is simply that the arrows of all three edges together should describe the same kind of rotation as the circular arrow on the 2-simplex. This can also be reduced to a purely combinatorial algorithm, and it makes sense in every dimension. For an n-simplex $\sigma = \{v_0, \ldots, v_n\}$, the ***k*th boundary face** $\partial_{(k)}\sigma$ of σ is the $(n-1)$-simplex whose vertices include all the $v_0, \ldots, v_n$ except v_k. Clearly if the vertices $v_0, \ldots, v_n$ come with an ordering, then the vertices of $\partial_{(k)}\sigma$ inherit an ordering from this, though here we

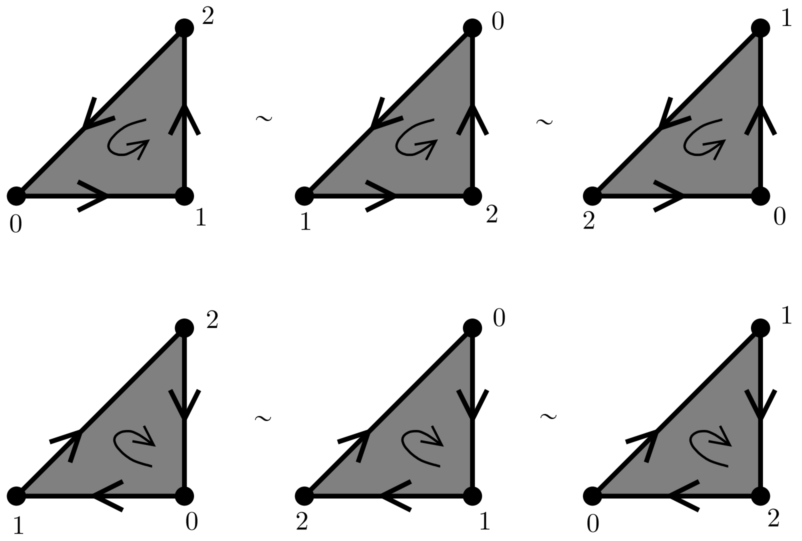


Figure 11. The six distinct orderings that define the two possible orientations of a 2-simplex.

have to be a bit careful because applying an even permutation to $v_0, \dots, v_n$ and then eliminating v_k may produce a sequence that differs from $v_0, \dots, v_{k-1}, v_{k+1}, \dots, v_n$ by an *odd* permutation. To get a well-defined orientation on $\partial_{(k)}\sigma$, one can instead do the following: notice that the sequence $v_0, \dots, v_k$ can be reordered as $v_k, v_0, \dots, v_{k-1}, v_{k+1}, \dots, v_n$ by a sequence of k flips. Permutations of this new sequence that fix the first object v_k are then equivalent to permutations of the vertices of $\partial_{(k)}\sigma$, so the even/odd parity of the permutation does not change if we remove v_k from the list. We must not forget however that in order to produce the list with v_k at the front, we performed k flips, meaning a permutation that is even if and only if k is even. This discussion implies that the following notion of boundary orientation is well defined.

Definition 20.8. Given an oriented n-simplex for $n \geqslant 2$ with vertices $v_0, \dots, v_n$ ordered accordingly, the induced **boundary orientation** of its kth boundary face $\partial_{(k)}\sigma$ is defined as the same ordering of its vertices (with v_k removed) if k is even, and otherwise it is defined by any odd permutation of this ordering. For $n = 1$, the boundary orientations are defined by assigning the sign $+1$ to $\partial_{(0)}\sigma = \{v_1\}$ and -1 to $\partial_{(1)}\sigma = \{v_0\}$.

You should now take a moment to stare again at Figure 11 and assure yourself that the boundary orientations indicated there are consistent with this definition.

Definition 20.9. An **oriented triangulation** of a closed surface Σ is a triangulation $\Sigma \cong |K|$ together with a choice of orientation for each 2-simplex in the complex K such that for every 1-simplex σ in K, the two induced boundary orientations that it inherits as a boundary face of two distinct 2-simplices are opposite.

The point of the condition on 1-simplices is to ensure that the orientations of any two neighboring 2-simplices are "compatible" in the sense that each of the circular arrows can be pushed continuously into the other. Figure 12 (left) shows an example of an oriented triangulation of $\mathbb{T}^2$. The arrows on 1-simplices in this picture are not meant to represent boundary orientations, but are just the usual indications of which 1-simplices on the boundary of the square should be glued

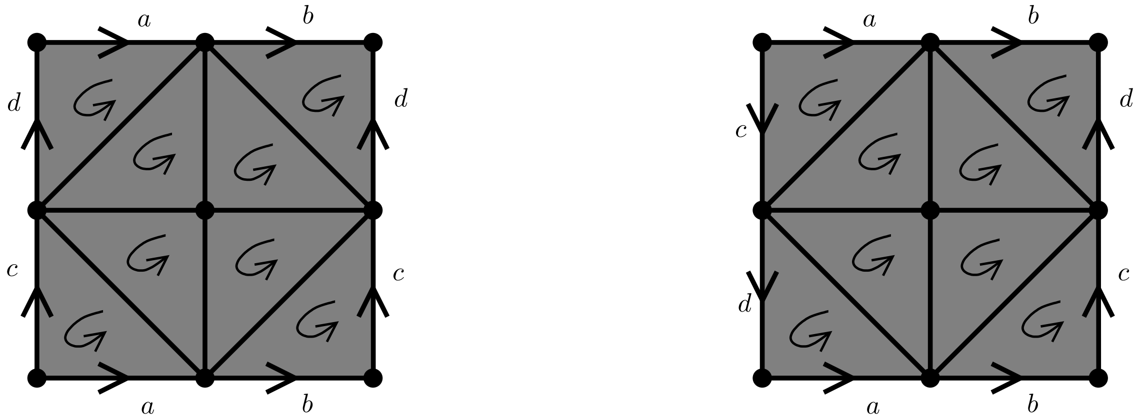


FIGURE 12. An oriented triangulation of the 2-torus (left) and a failed attempt to orient a triangulation of the Klein bottle (right).

together and how. We see in particular that the orientations indicated by these arrows on simplices c and d are the right boundary orientation on the right hand side but the wrong one on the left hand side. According to Definition 20.9, this is exactly what we want. Figure 12 (right) then shows what goes wrong if we try to do the same thing with a Klein bottle. If we imagine that this triangulation admits an orientation, then it will be represented by either clockwise or counterclockwise loops in each 2-simplex in the picture, all of them the same because they must induce opposite orientations on all the 1-dimensional boundary faces between them. In the picture they are all drawn counterclockwise. But notice that in both copies of each of the 1-simplices c and d, the arrow matches the induced boundary orientation, so this picture does not define a valid oriented triangulation. The next theorem implies in fact that no triangulation of the Klein bottle can be oriented.

THEOREM 20.10. *The following conditions are equivalent for any closed connected surface* Σ.

(1) Σ *is orientable.*
(2) Σ *admits an oriented triangulation.*
(3) Σ *does not contain any subset homeomorphic to the Möbius band.*

COROLLARY 20.11. *Every closed, connected and orientable surface is homeomorphic to* Σ_g *for some* $g \geqslant 0$. ☐

All of the ideas required for proving Theorem 20.10 have been discussed already, so let us merely sketch how they need to be put together. The equivalence of (1) and (2) is easy to understand by drawing small loops: clearly a choice of "counterclockwise loops" around points in the interior of any 2-simplex $\sigma \subset \Sigma$ determines a cyclic ordering of the vertices of that simplex, and conversely. Notice that this correspondence has a slightly non-obvious corollary: if some triangulation of Σ can be oriented, then so can all others. It should also be intuitively clear why (1) implies (3): if Σ contains a Möbius band, then no globally consistent notion of counterclockwise loops can be defined, since deforming it continuously along certain closed paths around the Möbius band would reverse it. For the converse, we can appeal to the classification of surfaces and observe that any surface Σ satisfying the third condition is homeomorphic to one of the surfaces Σ_g, which can be represented by a polygon with $4g$ sides. In the polygon picture, it is an easy exercise to construct an oriented triangulation for Σ_g. Alternatively, one can understand the relationship between (2) and (3) in terms of the presence of cross-caps or cross-handles in our proof of the classification of surfaces: the orientable surfaces are precisely those which can be constructed without any cross-caps or cross-handles, which turns out to work if and only if the 2-simplices can be assigned orientations for which the gluing maps between matching 1-simplices are orientation reversing.

Exercise 20.12. Construct an explicit oriented triangulation of Σ_g for each $g \geqslant 0$. Then, just for fun, count how many k-simplices it has for each $k = 0, 1, 2$. You will find that the number of 0-simplices minus the number of 1-simplices plus the number of 2-simplices is $2 - 2g$. (Someday next semester we'll discuss the Euler characteristic, and then you'll see why this is true.)

Exercise 20.13. In Exercise 14.13 we considered the space $\Sigma_{g,m}$, defined by cutting the interiors of $m \geqslant 0$ disjoint disks out of the oriented surface Σ_g of genus $g \geqslant 0$.

(a) Prove that every compact, orientable, connected surface with boundary is homeomorphic to $\Sigma_{g,m}$ for some values of $g, m \geqslant 0$.
Hint: If Σ is a compact 2-manifold, then $\partial\Sigma$ is a closed 1-manifold, and we classified all of the latter. With this knowledge, there is a cheap trick by which you can turn any compact surface with boundary into a closed surface, and then apply what you have learned about the classification of closed surfaces. Don't forget to keep track of orientations.

(b) Prove that $\Sigma_{g,m}$ is homeomorphic to $\Sigma_{h,n}$ if and only if $g = h$ and $m = n$.

This concludes our discussion of surfaces.

21. Higher homotopy, bordism, and simplicial homology

The rest of this semester's course will be about homology, but before defining it, I want to discuss some related ideas that should help motivate the definition. In some sense, all of the algebraic topological invariants we discuss in this course can be viewed as methods for "detecting holes" in a topological space. Let me start by describing a few concrete examples in which the fundamental group either does or does not succeed in this task.

Example 21.1. If we replace $\mathbb{R}^2$ with $\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2$, then the fundamental group changes from 0 to $\mathbb{Z}$, with the boundary of $\mathbb{D}^2$ representing a generator of $\pi_1(\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2)$, so this is one type of hole that π_1 detects very well.

Example 21.2. A 3-dimensional generalization of Example 21.1 is to replace $\mathbb{R}^3$ by $(\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2) \times \mathbb{R}$, which amounts to cutting the neighborhood of a line $\{0\} \times \mathbb{R} \subset \mathbb{R}^2 \times \mathbb{R}$ out of $\mathbb{R}^3$. Since the extra factor $\mathbb{R}$ is contractible, this example essentially admits a deformation retraction to the previous one, so we still find a generator of $\pi_1((\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2) \times \mathbb{R}) \cong \pi_1(\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2) \cong \mathbb{Z}$ which detects the removal of the tube $\mathring{\mathbb{D}}^2 \times \mathbb{R}$.

Example 21.3. A different type of generalization of Example 21.1 is to remove a 3-dimensional ball from $\mathbb{R}^3$, and here the fundamental group performs less well: $\pi_1(\mathbb{R}^3)$ is 0, and $\pi_1(\mathbb{R}^3 \backslash \mathring{\mathbb{D}}^3)$ is still zero since $\mathbb{R}^3 \backslash \mathring{\mathbb{D}}^3$ is homotopy equivalent to S^2 and the latter is simply connected. There clearly is a "hole" here, but π_1 does not see it.

Example 21.4. There are also examples in which π_1 seems to detect something other than a hole. Let $\Sigma_{g,m}$ denote the surface of genus g with m holes cut out, so Σ_2 is homeomorphic to a surface constructed by gluing together two copies of $\Sigma_{1,1}$ along their common boundary:

$$\Sigma_2 \cong \Sigma_{1,1} \cup_{\partial\Sigma_{1,1}} \Sigma_{1,1}.$$

Let $\gamma : S^1 \to \Sigma_2$ denote a loop parametrizing the common boundary of these copies of $\Sigma_{1,1}$. As we saw in Exercise 14.13, γ represents a nontrivial element in $\pi_1(\Sigma_2)$, though it is in the kernel of the natural homomorphism of $\pi_1(\Sigma_2)$ to its abelianization. The latter will turn out to be related to the following geometric observation: while γ cannot be extended to any map $\mathbb{D}^2 \to \Sigma_2$, it can be extended to a map on *some* surface with boundary S^1, e.g. it admits an extension to the inclusion $\Sigma_{1,1} \hookrightarrow \Sigma_2$. In this sense, there is no actual hole there for γ to detect; it is instead detecting a different phenomenon that has to do with the distinction between "disk-shaped" holes and "holes with genus".

I'm now going to start suggesting possible remedies for the drawbacks encountered in the last two examples. We will have to try a few times before we can point to the "right" remedy, but all of the objects we discuss along the way are also interesting and worthy of study.

Remedy 1: Higher homotopy groups. For any integer $k \geqslant 0$, fix a base point $t_0 \in S^k$ and associate to any pointed space (X, x_0) the set

$$\pi_k(X, x_0) = \left\{ f : (S^k, t_0) \to (X, x_0) \right\} / \underset{h+}{\sim},$$

where the equivalence relation $\underset{h+}{\sim}$ here means base-point preserving homotopy. This clearly reproduces the fundamental group when $k = 1$. When $k = 0$, $S^0 = \partial \mathbb{D}^1 = \{1, -1\}$ is a discrete space with two points, one of which must be the base point and is thus constrained to map to x_0, but the other can move freely within each path-component of X, so $\pi_0(X, x_0)$ is in bijective correspondence with the set of path-components of X. This set does not naturally have any group structure, though it does naturally have a "neutral" element, represented by the map that sends both points in S^0 to the base point x_0. It turns out that for $k \geqslant 2$, $\pi_k(X, x_0)$ can always be given the structure of an *abelian* group whose identity element is represented by the constant map

$$0 := [(S^k, t_0) \to (X, x_0) : t \mapsto x_0].$$

The precise definition of the group operation is a bit less obvious than for $k = 1$, so I will not go into it in this brief sketch. As with the fundamental group, one can show that $\pi_k(X, x_0)$ is independent of the base point up to isomorphism whenever X is path-connected, and it is also isomorphic for any two spaces that are homotopy equivalent. We will prove these statements next semester in *Topologie II*, but feel free to have a look at [**Hat02**, §4.1] if you can't bear to wait.

Here are a couple of things that can be proved about the higher homotopy groups using something resembling our present state of knowledge in this course:

EXAMPLE 21.5. The identity map $S^k \to S^k$ represents a nontrivial element of $\pi_k(S^k)$ for every $k \geqslant 1$. This follows from Exercise 19.14, which sketches the notion of the mod 2 mapping degree in order to show that every map $S^k \to S^k$ homotopic to the identity is surjective (and therefore nonconstant). More generally, one can use the integer-valued mapping degree for maps $S^k \to S^k$ to prove that $\pi_k(S^k) \cong \mathbb{Z}$, just like the case $k = 1$. A very nice account of this is given in [**Mil97**].

EXAMPLE 21.6. For every pair of integers $k, n \in \mathbb{N}$ with $n > k$, $\pi_k(S^n) = 0$. This follows easily from a general result in differential topology that allows us to approximate any continuous map between smooth manifolds by a smooth map for which any given point in the target space can be assumed to be a regular value. When $n > k$, the latter means that for any given $q \in S^n$ and a continuous map $f : S^k \to S^n$, we can approximate f with a map whose image does not contain q and is thus contained in $S^n \backslash \{q\} \cong \mathbb{R}^n$. The latter admits a deformation retraction to any point it contains, so composing the perturbed map $S^k \to S^n \backslash \{q\}$ with a deformation retraction of $S^n \backslash \{q\}$ to the base point gives a homotopy of f to the constant map.

Now here is the first piece of bad news about π_k: in general it is rather hard to compute. So hard, in fact, that the answers to certain basic questions about π_k remain unknown, e.g. one of the most popular open questions in modern topology is how to compute $\pi_k(S^n)$ in general when $k > n$. Various special cases are known, but the as-yet incomplete effort to extend these special cases to a general theorem has played a large role in motivating the development of modern homotopy theory.

We will need to have more and easier techniques at our disposal before we can discuss such things in earnest.

Remedy 2: Bordism groups. The higher homotopy groups do remedy one of the drawbacks of π_1 that I pointed out above: e.g. π_2 can be used to detect the hole in $\mathbb{R}^3 \backslash \mathring{\mathbb{D}}^3$ since, by homotopy invariance,

$$\pi_2(\mathbb{R}^3 \backslash \mathring{\mathbb{D}}^3) \cong \pi_2(S^2) \cong \mathbb{Z},$$

with the inclusion $S^2 \hookrightarrow \mathbb{R}^3 \backslash \mathring{\mathbb{D}}^3$ representing a generator. But there's another drawback here: while π_k can detect higher-dimensional holes, they are still holes of a fairly specific type which one might call "sphere-shaped" holes. What kind of hole is not sphere-shaped, you ask? Is there such a thing as a "torus-shaped" hole? How about this one:

Example 21.7. Let $X = S^1 \times \mathbb{R}^2$ and $X_0 = S^1 \times \mathring{\mathbb{D}}^2$, so $X \backslash X_0 = S^1 \times (\mathbb{R}^2 \backslash \mathring{\mathbb{D}}^2)$ admits a deformation retraction to $\partial \bar{X}_0 = S^1 \times S^1 = \mathbb{T}^2$. By homotopy invariance, we have $\pi_1(X) \cong \pi_1(S^1) \cong \mathbb{Z}$ and $\pi_1(X \backslash X_0) \cong \pi_1(\mathbb{T}^2) \cong \mathbb{Z}^2$, so π_1 does at least partly detect the removal of X_0 from X. But since $X \backslash X_0$ is homotopy equivalent to a surface, there is also an intrinsically 2-dimensional phenonomenon going on in this picture, and it seems natural to ask: does $X \backslash X_0$ contain any surface detecting the fact that X_0 has been removed from X? We can almost immediately give the following answer: if such a surface exists, it is *not* a sphere, in fact $\pi_2(X) = \pi_2(X \backslash X_0) = 0$. To see this, we can use the homotopy invariance of π_2: the spaces X and $X \backslash X_0$ are homotopy equivalent to S^1 and $\mathbb{T}^2$ respectively, so it suffices to prove $\pi_2(S^1) = \pi_2(\mathbb{T}^2) = 0$. Now observe that both S^1 and $\mathbb{T}^2$ are spaces whose universal covers ($\mathbb{R}$ and $\mathbb{R}^2$ respectively) happen to be contractible. In general, suppose $p : \widetilde{Y} \to Y$ denotes the universal cover of some reasonable space Y, and $\widetilde{Y}$ is contractible. Since S^2 is simply connected, any map $f : S^2 \to Y$ can be lifted to $\tilde{f} : S^2 \to \widetilde{Y}$, but the contractibility of $\widetilde{Y}$ then implies that $\tilde{f}$ is homotopic to a constant map. Composing that homotopy with $p : \widetilde{Y} \to Y$ gives a corresponding homotopy of $f = p \circ \tilde{f} : S^2 \to Y$ to a constant map, proving $\pi_2(Y) = 0$.

The preceding example is meant to provide motivation for a new invariant that might be able to detect holes that are not "sphere-shaped". The idea is to forget about the special roll played by spheres in the definition of π_k, but remember the fact that S^k is a closed k-dimensional manifold. Similarly, if M is a k-manifold, the homotopy relation for maps defined on M is defined in terms of maps on $I \times M$, which gives a special status to a very particular class of $(k+1)$-manifolds with boundary. Since we are now allowing arbitrary closed k-manifolds in place of spheres, it also seems natural to allow arbitrary compact $(k+1)$-manifolds with boundary for defining equivalence, instead of just manifolds of the form $I \times M$. Following this train of thought to its logical conclusion leads to *bordism theory.*[31]

For any space X and each integer $k \geqslant 0$, let

$$\Omega_k(X) := \{(M, f)\} / \sim,$$

---

[31] In the older literature, "bordism theory" was usually called "cobordism theory," and it is still common in most subfields of geometry and topology to refer to manifolds whose boundaries are disjoint unions of a given pair of closed manifolds as "cobordisms" instead of "bordisms". The elimination of the "co-" in "cobordism" is presumably motivated by the fact that bordism groups define a covariant functor instead of a contravariant functor, which makes it more analogous to *homology* than to *cohomology.* I promise you this footnote will make more sense after *Topologie II.*

where M is any closed (but not necessarily connected or nonempty)[32] k-manifold, $f : M \to X$ is a continuous map, and we write $(M_+, f_+) \sim (M_-, f_-)$ if and only if there exists a compact $(k+1)$-manifold W with $\partial W \cong M_- \amalg M_+$ and a map $F : W \to X$ such that $F|_{M_\pm} = f_\pm$. You should take a moment to think about why $\sim$ defines an equivalence relation. Any two pairs that are equivalent in this sense are said to be **bordant**, and the pair (W, F) is called a **bordism** between them.

EXAMPLE 21.8. $(M, f) \sim (M, g)$ whenever f and g are homotopic maps $M \to X$, as the homotopy $H : I \times M \to X$ defines a bordism $(I \times M, H)$.

EXAMPLE 21.9. Recall from Example 21.4 the loop $\gamma : S^1 \to \Sigma_2$ whose image separates Σ_2 into two pieces both homeomorphic to $\Sigma_{1,1}$. Either of the two inclusions $\Sigma_{1,1} \hookrightarrow \Sigma_2$ in this picture can be viewed as a bordism between (S^1, γ) and $(\varnothing, \cdot)$, where $\cdot$ denotes the unique map $\varnothing \to X$. Hence $[(S^1, \gamma)] = [(\varnothing, \cdot)] \in \Omega_1(\Sigma_2)$.

Since the manifolds representing elements of $\Omega_k(X)$ need not be connected, the disjoint union provides an obvious definition for a group operation on $\Omega_k(X)$. This operation is necessarily commutative since $X \amalg Y$ has a natural identification with $Y \amalg X$ for any two spaces X and Y. Now would be a good moment to mention the following notational convention: whenever a group G is known a priori to be abelian, we shall from now on denote the group operation in G as *addition* (with a "+" sign) rather than multiplication.

DEFINITION 21.10. We give $\Omega_k(X)$ the structure of an abelian group by defining

$$[(M_1, f_1)] + [(M_2, f_2)] := [(M_1 \amalg M_2, f_1 \amalg f_2)],$$

where $f_1 \amalg f_2 : M_1 \amalg M_2 \to X$ denotes the unique map whose restriction to $M_i \subset M_1 \amalg M_2$ is f_i for $i = 1, 2$. The identity element is

$$0 := [(\varnothing, \cdot)],$$

with $\cdot : \varnothing \to X$ denoting the unique map. The group $\Omega_k(X)$ is called the **k-dimensional unoriented bordism group** of X. We say that a pair (M, f) is **null-bordant** whenever $[(M, f)] = 0$, meaning there exists a compact $(k+1)$-manifold W with $\partial W \cong M$ and a map $F : W \to X$ with $F|_M = f$.

Referring back to Example 21.7, one can now show that the bordism class represented by the inclusion $\mathbb{T}^2 = \partial \bar{X}_0 \hookrightarrow X \backslash X_0$ is nontrivial in $\Omega_2(X \backslash X_0)$. One way to prove this uses the mod 2 mapping degree (cf. Exercise 19.14) for maps $f : \mathbb{T}^2 \to \mathbb{T}^2$: by an argument similar to the proof that $\deg_2(f)$ depends only on the homotopy class of f, one can show that $\deg(f) = 0$ whenever $(\mathbb{T}^2, f)$ is null-bordant. It follows that $[(\mathbb{T}^2, \mathrm{Id})] \neq 0 \in \Omega_2(\mathbb{T}^2)$ since $\deg_2(\mathrm{Id}) = 1$, and this element of $\Omega_2(\mathbb{T}^2)$ can be identified with the aforementioned inclusion using the homotopy equivalence between $\mathbb{T}^2$ and $X \backslash X_0$. In summary, Ω_2 does indeed detect "$\mathbb{T}^2$-shaped" holes.

The algebraic structure of $\Omega_k(X)$ is also extremely simple, one might even say too simple, in light of the following result saying that every element in $\Omega_k(X)$ is its own inverse:

PROPOSITION 21.11. *For every* $[(M, f)] \in \Omega_k(X)$, $[(M, f)] + [(M, f)] = 0$.

PROOF. Let $W = I \times M$ and $F : W \to X : (s, x) \mapsto f(x)$. Then $\partial W \cong \varnothing \amalg (M \amalg M)$ and $F|_{M \amalg M} = f \amalg f$, hence (W, F) is a bordism between $(M \amalg M, f \amalg f)$ and $(\varnothing, \cdot)$.[33] □

---

[32] Note that the empty set is a k-manifold for every $k \in \mathbb{Z}$. Look again at the definition of manifolds, and you will see that this is true.

[33] One of the slightly confusing things about $\Omega_k(X)$ is that there is always some ambiguity about how to split up the various connected components of ∂W into M_- and M_+. For the bordism in the proof of Prop. 21.11, one can equally well view it as a bordism between (M, f) and (M, f), but we are ignoring this because it does not give us any information beyond the fact that the bordism relation is reflexive.

One obtains a slightly more interesting algebraic structure by restricting to orientable manifolds and keeping track of orientations. Recall from the previous lecture that a manifold endowed with the extra structure of an orientation is called an *oriented manifold*; we will continue to denote such objects by single letters such as M, but you should keep in mind that they include slightly more data than just a set with its topology. If M is an oriented manifold, we shall denote by $-M$ the same manifold with its orientation reversed: this can always be defined by replacing each of the oriented charts on M by their compositions with an orientation-reversing homeomorphism $\mathbb{H}^n \to \mathbb{H}^n$ such as $(x_1, \ldots, x_{n-1}, x_n) \mapsto (x_1, \ldots, x_{n-1}, -x_n)$. Recall also from Remark 20.6 that any oriented manifold W with boundary determines a natural *boundary orientation* on ∂W. Whenever we write expressions like $\partial W \cong M$ in the context of oriented manifolds, we will always mean there is a homeomorphism $\partial W \to M$ that matches the given orientation of M to the boundary orientation of ∂W induced by the given orientation of W.

Definition 21.12. The **k-dimensional oriented bordism group** of X is[34]

$$\Omega_k^{\mathrm{SO}}(X) := \{(M, f)\}/\sim,$$

where M is a closed (but not necessarily connected or nonempty) oriented k-manifold, $f : M \to X$ is continuous, and the oriented bordism relation $(M_+, f_+) \sim (M_-, f_-)$ means that there exists a compact oriented $(k+1)$-manifold W and a map $F : W \to X$ such that

$$\partial W \cong -M_- \amalg M_+$$

and $F|_{M_\pm} = f_\pm$. The group operation on $\Omega_k^{\mathrm{SO}}(X)$ is defined via disjoint union as with $\Omega_k(X)$.

Proposition 21.11 is not true for oriented bordism groups: its proof fails due to the fact that the oriented boundary of $I \times M$ is $-M \amalg M$, not $M \amalg M$.

Let us compare both groups in the case $k = 0$. We claim that

$$\Omega_0(X) \cong \bigoplus_{\pi_0(X)} \mathbb{Z}_2,$$

while

$$\Omega_0^{\mathrm{SO}}(X) \cong \bigoplus_{\pi_0(X)} \mathbb{Z},$$

where $\pi_0(X)$ is an abbreviation for the set of path-components of X. For concreteness, consider a case where X has exactly three path-components $X_1, X_2, X_3 \subset X$, so the claim is that $\Omega_0(X) \cong \mathbb{Z}_2^3$ and $\Omega_0^{\mathrm{SO}}(X) \cong \mathbb{Z}^3$. An element of $\Omega_0(X)$ is an equivalence class of pairs (M, f), where M is a closed 0-manifold, i.e. a finite discrete set, and $f : M \to X$. Let us number the elements of M as $x_1, \ldots, x_N$, and suppose there are two elements that are mapped by f to the same path-component, say $f(x_1), f(x_2) \in X_1$. Then there exists a path $\gamma : I_{12} \to X$, where $I_{12} := I$, satisfying $\gamma(0) = f(x_1)$ and $\gamma(1) = f(x_2)$. Now define $W := I_{12} \amalg I_3 \amalg \ldots \amalg I_N$ where each I_j for $j = 3, \ldots, N$ is another copy of I, and decompose the boundary $\partial W = M_- \amalg M_+$ so that M_+ contains ∂I_{12} and $1 \in \partial I_j$ for every $j = 3, \ldots, N$, while M_- contains $0 \in \partial I_j$ for every $j = 3, \ldots, N$. Defining $F : W \to X$ such that $F|_{I_{12}} := \gamma$ and F sends I_j to the constant $f(x_j)$ for each $j = 3, \ldots, N$, we now have a bordism between (M, f) and (M', f') where $M' := M \backslash \{x_1, x_2\}$ and f' is the restriction of f. One can do this for any pair of points in M that are mapped to the same path-component, so that whenever (M, f) and (N, g) have the same number of points (mod 2) mapped into each path-component, there exists a bordism between them. Conversely, any bordism between two pairs (M, f) and (N, g) is of the form (W, F) where W is a compact 1-manifold with boundary,

---

[34] The "SO" in the notation $\Omega_k^{\mathrm{SO}}(X)$ stands for the group $\mathrm{SO}(k)$, the special orthogonal group. This has to do with the fact that $\mathrm{SO}(k)$ is precisely the subgroup of $\mathrm{O}(k)$ consisting of orthogonal transformations that are *orientation preserving*.

and by the classification of 1-manifolds, this can only mean a finite disjoint union of circles and compact intervals. Since each of these components individually can only be mapped into one of the path-components X_1, X_2, X_3 and each has either zero or two boundary points, it follows that for each $i = 1, 2, 3$, the number of points of M or N that are mapped into X_i can only differ by an even number. We have just proved the following: given $[(M, f)] \in \Omega_0(X)$, let $f_i \in \mathbb{Z}_2$ for $i = 1, 2, 3$ denote the number (mod 2) of points in M that f maps into X_i. Then

$$\Omega_0(X) \to \mathbb{Z}_2^3 : [(M, f)] \mapsto (f_1, f_2, f_3)$$

is an isomorphism.

To understand $\Omega_0^{\mathrm{SO}}(X)$, we need to keep in mind that an oriented 0-manifold M is not just a finite set of points, but it also comes with a map $\epsilon : M \to \{1, -1\}$ telling us which points are to be regarded as "positively oriented" as opposed to "negatively oriented" (cf. Definition 20.5). It is now no longer possible to cancel arbitrary pairs as in the unoriented case, but suppose $M = \{x_1, \ldots, x_N\}$ and f sends both x_1 and x_2 into X_1, and also that $\epsilon(x_1) = -1$ while $\epsilon(x_2) = +1$. We can again choose a path $\gamma : I_{12} \to X_1$ with $\gamma(0) = f(x_1)$ and $\gamma(1) = f(x_2)$, and define $W = I_{12} \amalg I_3 \amalg \ldots \amalg I_N$ and $F : W \to X$ as before. Before we can call (W, F) an oriented bordism, we need to specify the orientation of W. Let us assume I_{12} is oriented so that $\epsilon(1) = +1$ and $\epsilon(0) = -1$, while for $j = 3, \ldots, N$, orient I_j such that $\epsilon(1) = \epsilon(x_j)$ and $\epsilon(0) = -\epsilon(x_j)$. We now have $\partial W = -M' \amalg M$ where $M' = M \backslash \{x_1, x_2\}$ with the same orientations on the points $x_3, \ldots, x_N$, hence (W, F) is an oriented bordism between (M, f) and (M', f'). It is possible to construct such a bordism to eliminate any pair of points in M that have opposite signs and are mapped to the same path-component of X. Thus if we define $f_i \in \mathbb{Z}$ for each $i = 1, 2, 3$ by

$$f_i := \sum_{x \in f^{-1}(X_i)} \epsilon(x),$$

it follows that any two pairs (M, f) and (N, g) for which $f_i = g_i$ for every i must admit an oriented bordism. Conversely, the classification of 1-manifolds again implies that an arbitrary oriented bordism (W, F) between two pairs (M, f) and (N, g) is a map defined on a finite disjoint union of oriented intervals and circles, and since the two boundary points of an oriented interval I are always oriented with opposite signs, any component of W whose boundary lies entirely in one of M or $-N$ contributes zero to the counts defining the numbers f_i and g_i, while components that have one boundary point in M and one in $-N$ make the same contribution ± 1 to f_i and g_i. This proves that the map

$$\Omega_0^{\mathrm{SO}}(X) \to \mathbb{Z}^3 : [(M, f)] \mapsto (f_1, f_2, f_3)$$

is well defined and is also an isomorphism.

While computing the 0-dimensional bordism groups is not hard, we run into a serious (though interesting!) difficulty with the higher-dimensional bordism groups: they can be nontrivial even if X is only a one-point space. When $X = \{\mathrm{pt}\}$, we abbreviate

$$\Omega_k := \Omega_k(\{\mathrm{pt}\}), \qquad \Omega_k^{\mathrm{SO}} := \Omega_k^{\mathrm{SO}}(\{\mathrm{pt}\}),$$

and notice that since there is only one map from each manifold to $\{\mathrm{pt}\}$, the elements of Ω_k^{SO} are equivalence classes of oriented closed manifolds M where $M \sim N$ whenever $\partial W \cong -M \amalg N$ for some compact oriented manifold W; elements of Ω_k can be described in the same way after deleting the word "oriented" everywhere. In particular, we have $[M] = 0 \in \Omega_k$ if and only if M is homeomorphic to the boundary of some compact $(k+1)$-manifold. The question of whether a given manifold can be the boundary of another compact manifold is interesting, and the answer is often not obvious. For $k = 1$ it is not so hard: the classification of 1-manifolds implies that every bordism class $[M]$ in Ω_1 or Ω_1^{SO} is represented by a finite disjoint union of circles, and since

$S^1 = \partial\mathbb{D}^2$, all of these are (oriented) boundaries, hence

$$\Omega_1 = \Omega_1^{\mathrm{SO}} = 0.$$

It is similarly easy to see that all closed oriented surfaces are boundaries of compact oriented 3-manifolds: just take your favorite embedding of Σ_g into $\mathbb{R}^3$ and consider the region bounded by that embedded surface. For the oriented 3-dimensional case, we do not have any simple classification result to rely upon, but one can instead appeal to a standard (though not so trivial) result from low-dimensional topology known as the Dehn-Lickorish theorem, which can be interpreted as presenting arbitrary closed oriented 3-manifolds as boundaries of compact oriented 4-manifolds obtained by attaching "2-handles" to $\mathbb{D}^4$. We can therefore say

$$\Omega_2^{\mathrm{SO}} = \Omega_3^{\mathrm{SO}} = 0.$$

However, in the unoriented case there is already trouble in dimension two: it is known that there does not exist any compact 3-manifold whose boundary is homeomorphic to $\mathbb{RP}^2$. This can be proved using methods that we will cover in *Topologie II*, notably the Poincaré duality isomorphism between the homology and cohomology groups of closed manifolds. A similar argument implies that the complex counterpart of $\mathbb{RP}^2$, the complex projective space $\mathbb{CP}^2$, is a closed oriented 4-manifold that never occurs as the boundary of any compact oriented 5-manifold. This implies

$$[\mathbb{RP}^2] \neq 0 \in \Omega_2, \qquad \text{and} \qquad [\mathbb{CP}^2] \neq 0 \in \Omega_4^{\mathrm{SO}}.$$

This reveals that in general, the k-dimensional bordism groups of a one-point space contain a lot more information than one might expect: instead of just telling us something about the rather boring space $\{\mathrm{pt}\}$, they tell us something about the classification of closed k-manifolds, namely which ones can appear as boundaries of other compact manifolds and which ones cannot. That is an interesting question, and one that is very much worth studying at some point, but as with the higher homotopy groups, we will need to have a much wider range of simpler techniques at our disposal before we are equipped to tackle it.

Remedy 3: Simplicial homology (AKA "triangulated bordism"). The first version of homology theory that we will now discuss can be regarded as an attempt to capture much of the same information about X that is seen by the bordism groups $\Omega_n(X)$ and $\Omega_n^{\mathrm{SO}}(X)$, but without requiring us to know anything about the (generally quite hard) problem of classifying closed n-manifolds. The first idea is that instead of allowing arbitrary closed manifolds as domains, we consider manifolds with triangulations, so that all the data can be expressed in terms of simplices. The followup idea is that now that everything is expressed in terms of simplices, there is no need to mention manifolds at all.

Consider a simplicial complex $K = (V, S)$ with associated polyhedron $X := |K|$, and for each integer $n \geqslant 0$, let $S_{(n)} \subset S$ denote the set of n-simplices. As auxiliary data, we also fix an abelian group G, which in principle can be arbitrary, but for reasons related to the distinction between oriented and unoriented bordism, we will typically want to choose G to be either $\mathbb{Z}$ or $\mathbb{Z}_2$.

Definition 21.13. The group of **n-chains** in K (with **coefficients** in G) is the abelian group

$$C_n(K; G) := \bigoplus_{\sigma \in S_{(n)}} G,$$

whose elements can be written as finite sums $\sum_i a_i \sigma_i$ with $a_i \in G$ and $\sigma_i \in S_{(n)}$, with the group operation defined by

$$\sum_i a_i \sigma_i + \sum_i b_i \sigma_i = \sum_i (a_i + b_i)\sigma_i.$$

An n-chain is in some sense an abstract algebraic object, but if we choose $G = \mathbb{Z}$ and consider an n-chain $\sum_i a_i \sigma_i$ whose coefficients are all $a_i = \pm 1$, then you can picture the chain geometrically as the union of the n-simplices in X corresponding to each σ_i in the sum, with orientations determined by the signs a_i. These subsets are always compact, and if the particular set of n-simplices is chosen appropriately, then they will sometimes look like n-dimensional manifolds embedded in X. Our goal is now to single out a special class of n-chains that are analogous to *closed* n-dimensional manifolds embedded in X, i.e. the n-chains that have "empty boundary". This can be done by writing down an algebraic operation that describes the boundary of each individual simplex. To define this properly, we need to choose an orientation for every simplex in S; note that this has nothing intrinsically to do with oriented triangulations, as it is a completely arbitrary choice with no compatibility conditions required, so it can always be done. With this choice in place, for each $\sigma = \{v_0, \ldots, v_n\} \in S_{(n)}$, set

$$\partial \sigma := \sum_{k=0}^{n} \epsilon_k \partial_{(k)} \sigma \in C_{n-1}(K; \mathbb{Z}),$$

where as usual $\partial_{(k)} \sigma = \{v_0, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n\}$ denotes the kth boundary face of σ, and $\epsilon_k \in \{1, -1\}$ is defined to be $+1$ if the chosen orientation of the $(n-1)$-simplex $\partial_{(k)} \sigma$ matches the boundary orientation it inherits from σ (see Definition 20.8), and -1 if these two orientations are opposite. There is now a uniquely determined group homomorphism

$$\partial_n : C_n(K; G) \to C_{n-1}(K; G) : \sum_i a_i \sigma_i \mapsto \sum_i a_i (\partial \sigma_i),$$

where the multiplication of each coefficient $a_i \in G$ by a sign $\epsilon_k = \pm 1$ is defined in the obvious way as an element of G. (Notice that if $G = \mathbb{Z}_2$, the signs ϵ_k become irrelevant because every coefficient a_i then satisfies $a_i = -a_i$.) Strictly speaking, the definition above only makes sense for $n \geqslant 1$ since there are no (-1)-simplices; in light of this, we set

$$\partial_0 := 0.$$

We call the subgroup $\ker \partial_n \subset C_n(K; G)$ the group of n-**cycles**, or equivalently, the **closed** n-chains. The elements of the subgroup $\operatorname{im} \partial_{n+1} \subset C_n(K; G)$ are called **boundaries**.

Lemma 21.14. *$\partial_{n-1} \circ \partial_n = 0$ for all $n \in \mathbb{N}$.*

Proof. You should think of this as an algebraic or combinatorial expression of the geometric fact that the boundary of any n-manifold with boundary is always an $(n-1)$-manifold with *empty* boundary. On a more mundane level, the result holds due to cancelations, e.g. suppose A is an oriented 2-simplex whose oriented 1-dimensional boundary faces are denoted by a, b, c, giving

$$\partial_2 A = a + b + c.$$

Suppose further that the vertices of A are denoted by α, β, γ, all oriented with positive signs, but the arrow determined by the orientation of a points toward α and away from γ, while b points toward β and away from α, and c points toward γ but away from β. This gives the three relations

$$\partial_1 a = \alpha - \gamma, \quad \partial_1 b = \beta - \alpha, \quad \partial_1 c = \gamma - \beta,$$

thus $\partial_1 \circ \partial_2 A = \partial_1(a + b + c) = (\alpha - \gamma) + (\beta - \alpha) + (\gamma - \beta) = 0$. Similar cancelations occur in every dimension. □

Lemma 21.14 is often abbreviated with the formula

$$\partial^2 = 0,$$

and we will sometimes abbreviate $\partial := \partial_n$ when there is no chance of confusion. The formula implies in particular that $\operatorname{im}\partial_{n+1}$ is a subgroup of ∂_n for every $n \geqslant 0$. Since all these groups are abelian and subgroups are therefore normal, we can now consider quotients:

Definition 21.15. The nth **simplicial homology** group of the complex K (with coefficients in G) is

$$H_n^{\Delta}(K;G) := \ker\partial_n \big/ \operatorname{im}\partial_{n+1}.$$

It is worth comparing this definition to the bordism groups $\Omega_n(X)$ and $\Omega_n^{\mathrm{SO}}(X)$, as the extra layer of algebra involved in the definition of homology obscures a fairly direct analogy. Instead of closed n-manifolds M with maps $f : M \to X$, homology considers n-cycles, meaning formal linear combinations of n-simplices $c := \sum_i a_i\sigma_i$ with $\partial c = 0$. The bordism relation $(M_+, f_+) \sim (M_-, f_-)$ is now replaced by the conditition that two cycles $c, c' \in \ker\partial_n$ represent the same homology class $[c] = [c'] \in H_n^{\Delta}(K;G)$ if $c - c' \in \operatorname{im}\partial_{n+1}$, i.e. their difference is the boundary of an $(n+1)$-chain (analogous to a map defined on a compact $(n+1)$-manifold with boundary). When this holds, we say that the cycles c and c' are **homologous**. Finally, we will see that the distinction between $\Omega_n^{\mathrm{SO}}(X)$ and $\Omega_n(X)$ now corresponds to the distinction between $H_n^{\Delta}(K;\mathbb{Z})$ and $H_n^{\Delta}(K;\mathbb{Z}_2)$.

Let's compute an example. Figure 13 shows an oriented triangulation of $\mathbb{T}^2$ with eighteen 2-simplices, twenty-seven 1-simplices, and nine vertices labeled as follows:

$$\begin{aligned}
S_2 &= \{\sigma_1, \tau_1, \dots, \sigma_9, \tau_9\},\\
S_1 &= \{a_1, a_2, a_3, b_1, b_2, b_3, \dots, f_1, f_2, f_3, g_1, \dots, g_9\},\\
S_0 &= \{P_1, P_2, P_3, Q_1, Q_2, Q_3, R_1, R_2, R_3\}.
\end{aligned}$$

In addition to the orientations of the 2-simplices that come from this being an oriented triangulation, the figure shows (via arrows) an arbitrary choice of orientations for all 1-simplices, and we shall assume all the 0-simplices are oriented with a positive sign. One can now begin writing down relations such as

$$\partial\sigma_1 = g_1 - a_1 - d_3, \quad \partial\tau_1 = b_1 + e_3 - g_1, \quad \partial a_1 = P_2 - P_1$$

and so forth, but writing down all such relations would be rather tedious, so let us instead try to reason more geometrically. The computation of $H_0^{\Delta}(K;\mathbb{Z})$ is not hard in any case: all 0-chains are cycles since $\partial_0 = 0$, including the nine generators P_i, Q_i, R_i for $i = 1, 2, 3$, but all nine of them are also homologous to each other since any pair of them can be connected by a path of oriented 1-simplices leading from one to the other, e.g. $\partial a_1 = P_2 - P_1$ implies $[P_1] = [P_2]$, and $\partial e_3 = P_2 - R_2$ implies $[P_2] = [R_2]$. The result is

$$H_0^{\Delta}(K;\mathbb{Z}) \cong \mathbb{Z},$$

with a canonical generator represented by any of the vertices in the complex. Notice that this matches the oriented bordism group $\Omega_0^{\mathrm{SO}}(\mathbb{T}^2)$ since $\mathbb{T}^2$ is path-connected.

Let's look at the 1-cycles. There is a 1-cycle for every continuous loop we can find that follows a path through 1-simplices—we just have to insert minus signs wherever there is an arrow pointing the wrong way, in order to ensure the necessary cancelation of 0-simplices. For example, traversing the boundary of the lower-right square gives

$$\partial(a_3 + d_1 - c_3 - f_1) = 0,$$

so $a_3 + d_1 - c_3 - f_1$ is a 1-cycle, but not a very interesting one, since it is also the boundary of the region filled by the 2-simplices σ_9 and τ_9: in particular,

$$\partial(\sigma_9 + \tau_9) = (g_9 - c_3 - f_1) + (a_3 + d_1 - g_9) = a_3 + d_1 - c_3 - f_1,$$

hence $[a_3 + d_1 - c_3 - f_1] = 0 \in H_1^{\Delta}(K;\mathbb{Z})$. To find more interesting 1-cycles, it helps to remember what we already know about $\pi_1(\mathbb{T}^2) \cong \mathbb{Z}^2$. We can easily find two loops through 1-simplices that

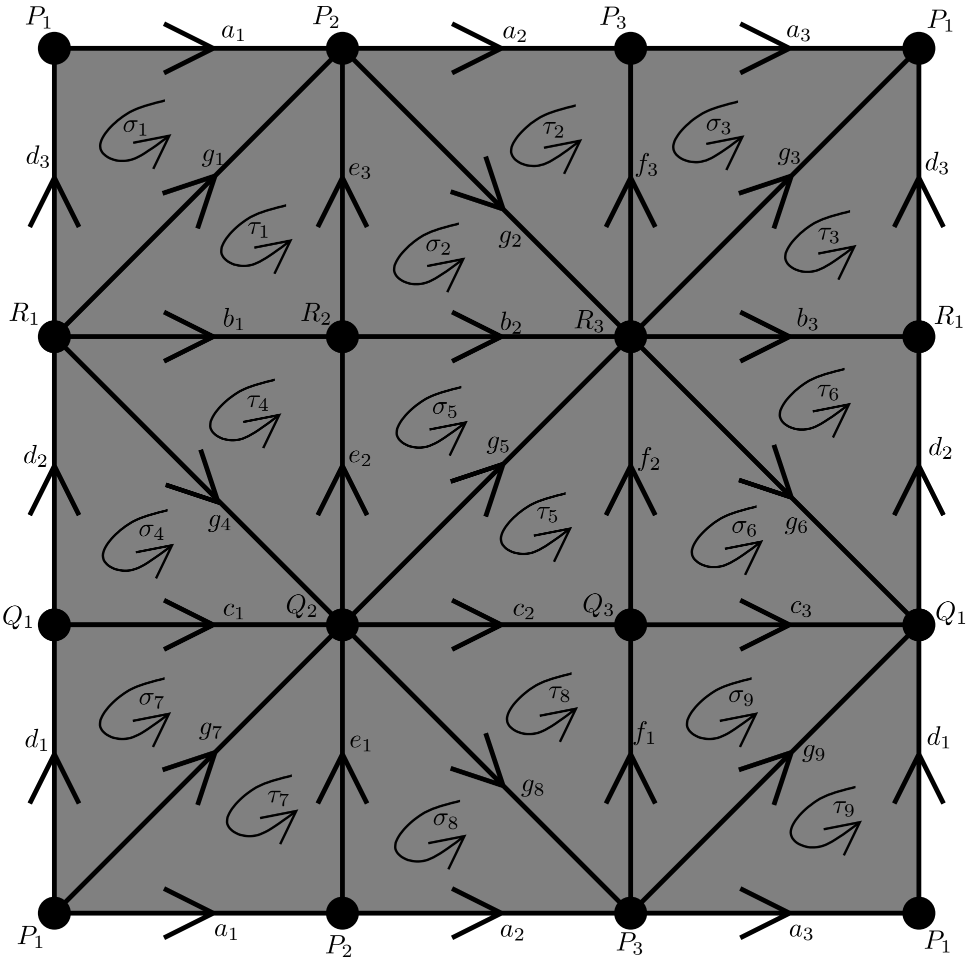


FIGURE 13. A simplicial complex with $|K| = \mathbb{T}^2$.

represent the two distinct generators of this fundamental group: one of them is $a_1 + a_2 + a_3$, and we easily see that

$$\partial(a_1 + a_2 + a_3) = (P_2 - P_1) + (P_3 - P_2) + (P_1 - P_3) = 0.$$

Another is $b_1 + b_2 + b_3$, but notice that the loops corresponding to these two 1-cycles are homotopic in $\mathbb{T}^2$, and relatedly, they form the boundary of the region filled by the six 2-simplices σ_i, τ_i for $i = 1, 2, 3$,

$$\partial(\sigma_1 + \sigma_2 + \sigma_3 + \tau_1 + \tau_2 + \tau_3) = (b_1 + b_2 + b_3) - (a_1 + a_2 + a_3),$$

implying $[a_1 + a_2 + a_3] = [b_1 + b_2 + b_3] \in H_1^{\Delta}(K; \mathbb{Z})$. Similar reasoning shows that $c_1 + c_2 + c_3$ is yet another 1-cycle representing the same homology class as both of these. One can show however that this homology class really is nontrivial, and it is not the only one: the other generator of $\pi_1(\mathbb{T}^2)$ corresponds to any of the three homologous 1-cycles $d_1 + d_2 + d_3$, $e_1 + e_2 + e_3$ or $f_1 + f_2 + f_3$. The end result is

$$H_1^{\Delta}(K; \mathbb{Z}) \cong \mathbb{Z}^2,$$

the same as the fundamental group.

As observed at the beginning of this lecture, the fact that $\mathbb{T}^2$ has a contractible universal cover implies that $\pi_2(\mathbb{T}^2) = 0$, so if there are any interesting 2-cycles in $\mathbb{T}^2$, they will not look like spheres. But if you think that $H_2(K; \mathbb{Z})$ should have something to do with the oriented bordism group $\Omega_2^{\mathrm{SO}}(\mathbb{T}^2)$, then there is a fairly obvious candidate for a 2-cycle in this picture: $\mathbb{T}^2$ itself is a closed oriented manifold, and the oriented triangulation we have chosen turns it into a 2-cycle:

$$\partial(\sigma_1 + \tau_1 + \ldots + \sigma_9 + \tau_9) = 0.$$

The point is that since the triangulation is oriented, writing down each individual term in this sum would produce a linear combination of 1-simplicies in which every 1-simplex in the complex appears exactly twice, but with opposite signs, thus adding up to 0. It should be easy to convince yourself that no nontrivial 2-chain that does not include all eighteen of the 2-simplices can ever be a cycle, as its boundary will have to include some 1-simplices that have nothing to cancel with. It follows easily that all 2-cycles in this complex are integer multiples of the one found above, and none of them are boundaries, since there are no 3-simplices, thus

$$H_2^{\Delta}(K; \mathbb{Z}) \cong \mathbb{Z}.$$

I can now state a theorem that is really rather amazing, though I'm sorry to say that we will not be able to prove it until next semester:

Theorem 21.16. *For any simplicial complex K, the simplicial homology groups $H_n^{\Delta}(K; G)$ depend (up to isomorphism) on the topological space $X = |K|$, i.e. the polyhedron of K, but not on the complex K itself.*

This theorem seems to have been known for quite a while before the reasons behind it were properly understood. At the dawn of homology theory, the subject had a very combinatorial flavor, and the use of triangulations as a tool for understanding manifolds proved to be very successful. A fairly natural strategy for proving Theorem 21.16 was formulated near the beginning of the 20th century and was based on a conjecture called the **Hauptvermutung**:[35] it claims essentially that any two triangulations of the same topological space can be turned into the same triangulation by a process of subdivision. Subdivision replaces each individual simplex σ with a triangulation by smaller simplices, so it makes the chain groups $C_n(K; G)$ much larger, but it is not too hard to show that the homology resulting from these enlarged chain groups is isomorphic to the original, hence if the Hauptvermutung is true, Theorem 21.16 follows. The only trouble is that the Hauptvermutung is false, as was discovered in the 1960's; moreover, we now also know examples of closed topological manifolds that cannot be triangulated at all, so that simplicial complexes do not provide the ideal framework for understanding manifolds in general. But in the mean time, the mathematical community discovered much better ways of proving Theorem 21.16, namely by defining another invariant for arbitrary topological spaces X that manifestly only depends on the topology of X without any auxiliary structure, but also can be shown to match simplicial homology whenever X is a polyhedron. That invariant is singular homology, and it will be our topic for the rest of this semester.

22. Singular homology

So here's the challenge: how do we define a topological invariant that captures the same information as simplicial homology, but without ever referring to a simplicial complex? The answer to this turns out to be fairly simple, but speaking for myself, the first time I heard it, I thought it sounded crazy. There seemed to be no way that one could ever compute such a thing, or if one could, then it was hard to imagine what geometric insight would be gained from the computation.

---

[35] This is what the conjecture was called in English—one does not translate the word *Hauptvermutung*.

I've been leading up to this definition gradually over the last few lectures in order to give you some intuition about what kind of invariant we are looking for and why. The hope is that, equipped with this intuition, your first reaction to seeing the definition of singular homology might be that it has a fighting chance of answering some question you actually care about.

It will be convenient to first establish some basic principles of the subject known as *homological algebra*. We have already seen an example of the first definition in our discussion of simplicial homology.

Definition 22.1. A ($\mathbb{Z}$-graded) **chain complex** (*Kettenkomplex*) of abelian groups (C_*, ∂) consists of a sequence $\{C_n\}_{n \in \mathbb{Z}}$ of abelian groups together with homomorphisms $\partial_n : C_n \to C_{n-1}$ for each $n \in \mathbb{Z}$ such that $\partial_{n-1} \circ \partial_n : C_n \to C_{n-2}$ is the trivial homomorphism for every n.

We sometimes denote the direct sum of all the chain groups C_n in a chain complex by

$$C_* := \bigoplus_{n \in \mathbb{Z}} C_n,$$

whose elements can all be written as finite sums $\sum_i a_i$ with $a_i \in C_{n_i}$ for some integers $n_i \in \mathbb{Z}$. An element $x \in C_*$ is said to have **degree** (*Grad*) n if $x \in C_n$. The individual homomorphisms $\partial_n : C_n \to C_{n-1}$ extend uniquely to a homomorphism $\partial : C_* \to C_*$ which has **degree** -1, meaning it maps elements of any given degree to elements of one degree less. We sometimes indicate this by abusing notation and writing

$$\partial : C_* \to C_{*-1}.$$

The collection of relations $\partial_{n-1} \circ \partial_n = 0$ for all n can now be abbreviated by the single relation

$$\partial^2 = 0,$$

which is equivalent to the condition that $\operatorname{im} \partial_{n+1} \subset \ker \partial_n$ for every n. We call ∂ the **boundary map** (*Randoperator*) in the complex. Elements in $\ker \partial \subset C_*$ are called **cycles** (*Zykel*), while elements in $\operatorname{im} \partial \subset C_*$ are called **boundaries** (*Ränder*).

Definition 22.2. The **homology** (*Homologie*) of a chain complex (C_*, ∂) is the sequence of abelian groups

$$H_n(C_*, \partial) := \ker \partial_n \big/ \operatorname{im} \partial_{n+1}$$

for $n \in \mathbb{Z}$. We sometimes denote

$$H_*(C_*, \partial) := \bigoplus_{n \in \mathbb{Z}} H_n(C_*, \partial),$$

which makes $H_*(C_*, \partial)$ a $\mathbb{Z}$-graded abelian group.

Every element of $H_n(C_*, \partial)$ can be written as an equivalence class $[c]$ for some n-cycle $c \in \ker \partial_n$, and we call $[c]$ the **homology class** (*Homologieklasse*) represented by c. Two cycles $a, b \in \ker \partial_n$ are called **homologous** (*homolog*) if $[a] = [b] \in H_n(C_*, \partial)$, meaning $a - b \in \operatorname{im} \partial_{n+1}$.

Remark 22.3. For the examples of chain complexes (C_*, ∂) we consider in this course, C_n is always the trivial group for $n < 0$, mainly because the degree n typically corresponds to a geometric dimension and dimensions cannot be negative. But there is no need to assume this in the general algebraic definitions. In other settings, there are plenty of interesting examples of chain complexes that have nontrivial elements of negative degree.

The next definition will be needed when we want to show that continuous maps between topological spaces induce homomorphisms of their singular homology groups.

Definition 22.4. Given two chain complexes (A_*, ∂^A) and (B_*, ∂^B), a **chain map** (*Kettenabbildung*) from (A_*, ∂^A) to (B_*, ∂^B) is a sequence of homomorphisms $f_n : A_n \to B_n$ for $n \in \mathbb{Z}$ such that the following diagram commutes:

$$\begin{tikzcd}
\dots \arrow[r] & A_{n+1} \arrow[r, "\partial^A_{n+1}"] \arrow[d, "f_{n+1}"] & A_n \arrow[r, "\partial^A_n"] \arrow[d, "f_n"] & A_{n-1} \arrow[r, "\partial^A_{n-1}"] \arrow[d, "f_{n-1}"] & \dots \\
\dots \arrow[r] & B_{n+1} \arrow[r, "\partial^B_{n+1}"] & B_n \arrow[r, "\partial^B_n"] & B_{n-1} \arrow[r, "\partial^B_{n-1}"] & \dots
\end{tikzcd} \tag{22.1}$$

In other words, a chain map is a homomorphism $f : A_* \to B_*$ of degree zero satisfying $\partial^B \circ f = f \circ \partial^A$.

Proposition 22.5. *Any chain map $f : (A_*, \partial^A) \to (B_*, \partial^B)$ determines homomorphisms $f_* : H_n(A_*, \partial^A) \to H_n(B_*, \partial^B)$ for every $n \in \mathbb{Z}$ via the formula*

$$f_*[a] := [f(a)].$$

Proof. There are two things to prove: first, that whenever $a \in A_n$ is a cycle, so is $f(a) \in B_n$. This is clear since $\partial^A a = 0$ implies $\partial^B(f(a)) = f(\partial^A a) = 0$ by the chain map condition. Second, we need to know that f maps boundaries to boundaries, so that it descends to a well-defined homomorphism $\ker \partial^A_n / \operatorname{im} \partial^A_{n+1} \to \ker \partial^B_n / \operatorname{im} \partial^B_{n+1}$. This is equally clear, since $a = \partial^A x$ implies $f(a) = f(\partial^A x) = \partial^B f(x)$. □

With these algebraic preliminaries out of the way, we now proceed to define the chain complex of singular homology. As in simplicial homology, we fix an arbitrary abelian group G as auxiliary data, called the **coefficient group**; in practice it will usually be either $\mathbb{Z}$ or $\mathbb{Z}_2$, occasionally $\mathbb{Q}$. Recall that for integers $n \geqslant 0$, the **standard n-simplex** is the set

$$\Delta^n = \{(t_0, \dots, t_n) \in I^{n+1} \mid t_0 + \dots + t_n = 1\}.$$

For each $k = 0, \dots, n$, the **kth boundary face** of Δ^n is the subset

$$\partial_{(k)}\Delta^n := \{t_k = 0\} \subset \Delta^n,$$

which is canonically homeomorphic to Δ^{n-1} via the map

$$\partial_{(k)}\Delta^n \to \Delta^{n-1} : (t_0, \dots, t_{k-1}, 0, t_{k+1}, \dots, t_n) \mapsto (t_0, \dots, t_{k-1}, t_{k+1}, \dots, t_n). \tag{22.2}$$

Definition 22.6. Given a topological space X, a **singular n-simplex** in X is a continuous map $\sigma : \Delta^n \to X$.

Let $\mathcal{K}_n(X)$ denote the set of all singular n-simplices in X, and define the **singular n-chain group** with coefficients in G by

$$C_n(X; G) = \bigoplus_{\sigma \in \mathcal{K}_n(X)} G.$$

Note that this definition also makes sense for $n < 0$ if we agree that $\mathcal{K}_n(X)$ is then empty since there is no such thing as a simplex of negative dimension, hence the groups $C_n(X; G)$ are trivial in these cases. In general, elements in $C_n(X; G)$ can be written as finite sums $\Sigma_i a_i \sigma_i$ where $a_i \in G$ and $\sigma_i \in \mathcal{K}_n(X)$. This clearly looks similar to the simplicial chain groups, but if you're paying attention properly, you may at this point be feeling nervous about the fact that $C_n(X; G)$ is a *bloody enormous* group: algebraically it is very simple, but the set $\mathcal{K}_n(X)$ that generates it is usually uncountably infinite. It's probably even larger than you are imagining, because a singular n-simplex is not just a "simplex-shaped" subset of X, but it is also the parametrization of that subset, so any two distinct parametrizations $\sigma : \Delta^n \to X$, even if they have exactly the same image, define different elements of $\mathcal{K}_n(X)$ and thus different generators of $C_n(X;G)$.[36] If this makes you nervous, then you are right to feel nervous: it is a minor miracle that we will eventually be able to extract useful and computable information from groups as large as $C_n(X;G)$. You will see.

The next step is to define a boundary map $C_n(X;G) \to C_{n-1}(X;G)$. As in simplicial homology, this is done by writing a formula for $\partial\sigma$ for each generator $\sigma \in \mathcal{K}_n(X)$, and the formula follows the same orientation convention that we saw in our discussion of oriented triangulations, cf. Definition 20.8: set

$$\partial\sigma := \sum_{k=0}^{n} (-1)^k \left(\sigma|_{\partial_{(k)}\Delta^n}\right) \in C_{n-1}(X;\mathbb{Z}),$$

where each $\sigma|_{\partial_{(k)}\Delta^n}$ is regarded as a singular $(n-1)$-simplex using the identification $\partial_{(k)}\Delta^n = \Delta^{n-1}$ from (22.2).

This uniquely determines a homomorphism

$$\partial : C_n(X;G) \to C_{n-1}(X;G) : \sum_i a_i\sigma_i \mapsto \sum_i a_i \; \partial\sigma_i,$$

and the usual cancelation phenomenon implies:

Lemma 22.7. $\partial^2 = 0$. ☐

The **nth singular homology group** (*singuläre Homologiegruppe*) with coefficients in G is now defined by

$$H_n(X;G) := H_n\left(C_*(X;G), \partial\right).$$

In the case $G = \mathbb{Z}$, this is often abbreviated by

$$H_n(X) := H_n(X;\mathbb{Z}).$$

The direct sum of these groups for all n is denoted by $H_*(X;G)$, though informally, this notation is also sometimes used with the symbol "$*$" acting as an integer-valued variable just like n.

I encourage you to compare the following result with our computation of the bordism groups $\Omega_0(X)$ and $\Omega_0^{\mathrm{SO}}(X)$ in Lecture 21.

Proposition 22.8. *For any space X and any coefficient group G, $H_0(X;G) \cong \bigoplus_{\pi_0(X)} G$, i.e. it is a direct sum of copies of G for every path-component of X.*

Proof. Since Δ^0 is a one-point space, the set $\mathcal{K}_0(X)$ of singular 0-simplices $\sigma : \Delta^0 \to X$ can be identified naturally with X, and we shall write 0-chains accordingly as finite sums $\sum_i a_i x_i$ with $a_i \in G$ and $x_i \in X$. Similarly, Δ^1 is homeomorphic to the unit interval $I = [0,1]$, and if we choose a homeomorphism $[0,1] \to \Delta^1$ sending 1 to $\partial_{(0)}\Delta^1$ and 0 to $\partial_{(1)}\Delta^1$, we can think of each $\sigma \in \mathcal{K}_1(X)$ as a path $\sigma : I \to X$ and write the boundary operator as

$$\partial\sigma = \sigma(1) - \sigma(0) \in C_0(X;\mathbb{Z}).$$

Since there are no (-1)-chains, every $a \in G$ and $x \in X$ then define a 0-cycle $ax \in C_0(X;G)$, but ax and ay are homologous whenever x and y belong to the same path-component since then any path $\sigma : I \to X$ from x to y gives $\partial(a\sigma) = ay - ax$. Choosing a point x_α in each path-component X_α, we can now say that every 0-cycle is homologous to a unique 0-cycle of the form $\sum_\alpha c_\alpha x_\alpha$, where the sum ranges over all the path-components of X but only finitely many of the coefficients $c_\alpha \in G$ are nonzero. If two cycles of this form are homologous, then they differ by the boundary of a 1-chain, which is a finite linear combination of paths, and since each path is confined to a single

---

[36] The word "singular" in this context refers to the fact that there is no condition beyond continuity required for the maps $\sigma : \Delta^n \to X$, i.e. they need not be injective, nor differentiable (even if X happens to be a smooth manifold), and so their images might not look "simplex-shaped" at all, but could instead be full of singularities.

path-component and has two end points with opposite orientations, the conclusion is that both 0-cycles have the same coefficients. □

The next result is a straightforward exercise based on the definitions, and you should also compare it with our previous discussion of the bordism groups of a point, if only to observe that the result is very different: while bordism groups require some information about the classification of manifolds which has nothing to do with the one-point space, the singular homology of $\{\mathrm{pt}\}$ is much simpler.

Exercise 22.9. Show that for the 1-point space $\{\mathrm{pt}\}$ and any coefficient group G, singular homology satisfies

$$H_n(\{\mathrm{pt}\}; G) \cong \begin{cases} G & \text{for } n = 0, \\ 0 & \text{for } n \neq 0. \end{cases}$$

Hint: For each integer $n \geqslant 0$, there is exactly one singular n-simplex $\Delta^n \to \{\mathrm{pt}\}$, so the chain groups $C_n(\{\mathrm{pt}\}; G)$ are all naturally isomorphic to G. What is $\partial : C_n(\{\mathrm{pt}\}; G) \to C_{n-1}(\{\mathrm{pt}\}; G)$?

Let us discuss the group $H_1(X; \mathbb{Z})$ for an arbitrary space X. As noted above in our proof of Proposition 22.8, Δ^1 is homeomorphic to the interval I, thus there is a bijection

$$\{\text{paths } I \to X\} \leftrightarrow \mathcal{K}_1(X) \tag{22.3}$$

which identifies each path γ with a singular 1-simplex (denoted by the same symbol) such that, under the canonical identification of $\mathcal{K}_0(X)$ with X,

$$\partial\gamma = \gamma(1) - \gamma(0).$$

Notice in particular that if γ is a loop, then it also defines a 1-cycle. More generally, let us write elements of $C_1(X; \mathbb{Z})$ as finite sums $\sum_i m_i\gamma_i$ where $m_i \in \mathbb{Z}$ and the γ_i are understood as singular 1-simplices via the above bijection, so

$$\partial \sum_i m_i\gamma_i = \sum_i m_i \left(\gamma_i(1) - \gamma_i(0)\right) \in C_0(X; \mathbb{Z}).$$

Now observe that since the coefficients m_i are integers, we are free to assume they are all ± 1 at the cost of allowing repeats in the finite list of paths γ_i. It will then be convenient to think of $-\gamma_i$ as the reversed path γ_i^{-1}, which makes sense if you look at the boundary formula since

$$\partial(-\gamma_i) = -(\gamma_i(1) - \gamma_i(0)) = \gamma_i(0) - \gamma_i(1) = \gamma_i^{-1}(1) - \gamma_i^{-1}(0) = \partial(\gamma_i^{-1}).$$

Thinking in these terms and continuing to assume $m_i = \pm 1$, $\sum_i m_i\gamma_i$ will now be a cycle if and only if the finite set of paths $\gamma_i^{m_i}$ can be arranged in some order so that they form a loop, i.e. each can be concatenated with the next in the list, and the last can be concatenated with the first. This is precisely what is needed in order to ensure that every 0-simplex in $\partial \sum_i m_i\gamma_i$ cancels out. This suggests a relationship between $H_1(X; \mathbb{Z})$ and $\pi_1(X)$, but notice that there is some ambiguity in the correspondence: in general there may be multiple ways that the paths $\gamma_i^{m_i}$ can be ordered to produce a loop, and different loops produced in this way need not always be homotopic to each other. In fact, one should not expect $H_1(X; \mathbb{Z})$ and $\pi_1(X)$ to be the same, since $H_1(X; \mathbb{Z})$ is abelian by definition, but $\pi_1(X)$ usually is not. It turns out that the next best thing is true.

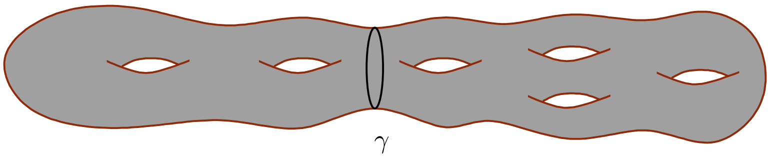
Theorem 22.10. *For any path-connected space X with base point $x_0 \in X$, the bijection (22.3) determines a group homomorphism*

$$h : \pi_1(X, x_0) \to H_1(X; \mathbb{Z})$$

which descends to an isomorphism of the abelianization $\pi_1(X, x_0)/[\pi_1(X, x_0), \pi_1(X, x_0)]$ to $H_1(X; \mathbb{Z})$.

We say that a cycle $c \in C_*(X;G)$ is **nullhomologous** if $[c] = 0 \in H_*(X;G)$, or equivalently, c is a boundary. According to the discussion above, every loop $\gamma : I \to X$ with $\gamma(0) = \gamma(1) = x_0$ can be viewed as a 1-cycle, and that cycle is nullhomologous if and only if $[\gamma]$ belongs to the commutator subgroup of $\pi_1(X, x_0)$.

Example 22.11. Recall from Exercise 14.13 the embedded loop $\gamma : S^1 \to \Sigma_g$ for $g \geqslant 2$ whose image separates Σ_g into two surfaces of genus $h \geqslant 1$ and $k \geqslant 1$ respectively with one boundary component each:



We computed in that exercise that $[\gamma]$ is a nontrivial element of the commutator subgroup of $\pi_1(\Sigma_g)$, thus by Theorem 22.10, γ represents the trivial class in $H_1(\Sigma_g;\mathbb{Z})$. This should not be surprising, since γ also parametrizes the boundary of a compact oriented submanifold of Σ_g, e.g. for this same reason, γ also represents the trivial bordism class in $\Omega_1^{\mathrm{SO}}(\Sigma_g)$. One can find an explicit 2-chain whose boundary is γ by decomposing the surface $\Sigma_{h,1}$ into 2-simplices so as to reinterpret the inclusion $\Sigma_{h,1} \hookrightarrow \Sigma_g$ as a linear combination of singular 2-simplices in Σ_g.

The proof of Theorem 22.10 is not trivial, but it is simple enough to leave as a guided homework problem (see Exercise 22.12 below). The homomorphism $h : \pi_1(X) \to H_1(X;\mathbb{Z})$ is called the **Hurewicz map**. There exists a similar Hurewicz homomorphism $\pi_k(X) \to H_k(X;\mathbb{Z})$ for every $k \geqslant 1$, which we will discuss near the end of *Topologie II* if time permits. Note that for $k \geqslant 2$, $\pi_k(X)$ is always abelian, so it is reasonable in those cases to hope that the Hurewicz map might be an honest isomorphism. A result called Hurewicz's theorem gives conditions under which this turns out to hold, thus providing a nice way to compute higher homotopy groups in some cases since, as we will see, computing homology is generally easier. But there are also simple examples in which $\pi_k(X)$ and $H_k(X;\mathbb{Z})$ are totally different. We saw for instance in the previous lecture that $\pi_2(\mathbb{T}^2) = 0$ due to the lifting theorem, but one can use any oriented triangulation of $\mathbb{T}^2$ to produce a singular 2-cycle that can be shown to be nontrivial in $H_2(\mathbb{T}^2;\mathbb{Z})$. Homology classes in the image of the Hurewicz map are sometimes called *spherical* homology classes. The example of $\mathbb{T}^2$ shows that for $n \geqslant 2$, one cannot generally expect all classes in $H_n(X;\mathbb{Z})$ to be spherical.

Exercise 22.12. Let us prove Theorem 22.10. Assume X is a path-connected space, fix $x_0 \in X$ and abbreviate $\pi_1(X) := \pi_1(X, x_0)$, so elements of $\pi_1(X)$ are represented by paths $\gamma : I \to X$ with $\gamma(0) = \gamma(1) = x_0$. Identifying the standard 1-simplex

$$\Delta^1 := \left\{(t_0, t_1) \in \mathbb{R}^2 \;\middle|\; t_0 + t_1 = 1,\ t_0, t_1 \geqslant 0\right\}$$

with $I := [0,1]$ via the homeomorphism $\Delta^1 \to I : (t_0, t_1) \mapsto t_1$, every path $\gamma : I \to X$ corresponds to a singular 1-simplex $\Delta^1 \to X$, which we shall denote by $\tilde{h}(\gamma)$ and regard as an element of the singular 1-chain group $C_1(X;\mathbb{Z})$. Show that $\tilde{h}$ has each of the following properties:

(a) If $\gamma : I \to X$ satisfies $\gamma(0) = \gamma(1)$, then $\partial\tilde{h}(\gamma) = 0$.
(b) For any constant path $e : I \to X$, $\tilde{h}(e) = \partial\sigma$ for some singular 2-simplex $\sigma : \Delta^2 \to X$.
(c) For any paths $\alpha, \beta : I \to X$ with $\alpha(1) = \beta(0)$, the concatenated path $\alpha \cdot \beta : I \to X$ satisfies $\tilde{h}(\alpha) + \tilde{h}(\beta) - \tilde{h}(\alpha \cdot \beta) = \partial\sigma$ for some singular 2-simplex $\sigma : \Delta^2 \to X$.
Hint: Imagine a triangle whose three edges are mapped to X via the paths α, β and $\alpha \cdot \beta$. Can you extend this map continuously over the rest of the triangle?

(d) If $\alpha, \beta : I \to X$ are two paths that are homotopic with fixed end points, then $\tilde{h}(\alpha) - \tilde{h}(\beta) = \partial f$ for some singular 2-chain $f \in C_2(X; \mathbb{Z})$.
Hint: If you draw a square representing a homotopy between α and β, you can decompose this square into two triangles.

(e) Applying $\tilde{h}$ to paths that begin and end at the base point x_0, deduce that $\tilde{h}$ determines a group homomorphism $h : \pi_1(X) \to H_1(X; \mathbb{Z}) : [\gamma] \mapsto [\tilde{h}(\gamma)]$.

We call $h : \pi_1(X) \to H_1(X; \mathbb{Z})$ the **Hurewicz homomorphism**. Notice that since $H_1(X; \mathbb{Z})$ is abelian, $\ker h$ automatically contains the commutator subgroup $[\pi_1(X), \pi_1(X)] \subset \pi(X)$ (see Exercise 12.21), thus h descends to a homomorphism on the abelianization of $\pi_1(X)$,

$$\Phi : \pi_1(X) \big/ [\pi_1(X), \pi_1(X)] \to H_1(X; \mathbb{Z}).$$

We will now show that this is an isomorphism by writing down its inverse. For each point $p \in X$, choose arbitrarily a path $\omega_p : I \to X$ from x_0 to p, and choose ω_{x_0} in particular to be the constant path. Regarding singular 1-simplices $\sigma : \Delta^1 \to X$ as paths $\sigma : I \to X$ under the usual identification of I with Δ^1, we can then associate to every singular 1-simplex $\sigma \in C_1(X; \mathbb{Z})$ a concatenated path

$$\tilde{\Psi}(\sigma) := \omega_{\sigma(0)} \cdot \sigma \cdot \omega_{\sigma(1)}^{-1} : I \to X$$

which begins and ends at the base point x_0, hence $\tilde{\Psi}(\sigma)$ represents an element of $\pi_1(X)$. Let $\Psi(\sigma)$ denote the equivalence class represented by $\tilde{\Psi}(\sigma)$ in the abelianization $\pi_1(X)/[\pi_1(X), \pi_1(X)]$. This uniquely determines a homomorphism[37]

$$\Psi : C_1(X; \mathbb{Z}) \to \pi_1(X)/[\pi_1(X), \pi_1(X)] : \sum_i m_i \sigma_i \mapsto \sum_i m_i \Psi(\sigma_i).$$

(f) Show that $\Psi(\partial\sigma) = 0$ for every singular 2-simplex $\sigma : \Delta^2 \to X$, and deduce that Ψ descends to a homomorphism $\Psi : H_1(X; \mathbb{Z}) \to \pi_1(X)/[\pi_1(X), \pi_1(X)]$.

(g) Show that $\Psi \circ \Phi$ and $\Phi \circ \Psi$ are both the identity map.

(h) For a closed surface Σ_g of genus $g \geqslant 2$, find an example of a nontrivial element in the kernel of the Hurewicz homomorphism $\pi_1(\Sigma_g) \to H_1(\Sigma_g)$. *Hint: See Exercise 14.13.*

23. Relative homology and long exact sequences

The above results for $H_0(X; G)$ and $H_1(X; \mathbb{Z})$ provide some evidence that in spite of being defined as quotients of groups with uncountably many generators, the singular homology groups $H_n(X; G)$ might turn out to be computable more often than we'd expect. In this lecture we'll introduce a powerful computational tool that is also a fundamental concept in homological algebra. But before that, let us clarify in what sense singular homology is a topological invariant.

Lemma 23.1. *Every continuous map $f : X \to Y$ determines a chain map $f_* : C_*(X; G) \to C_*(Y; G)$ via the formula $f_*\sigma := f \circ \sigma$ for singular n-simplices $\sigma : \Delta^n \to X$.*

Proof. It is straightforward to check that $\partial(f_*\sigma) = f_*(\partial\sigma) \in C_{n-1}(Y; \mathbb{Z})$ for all $\sigma : \Delta^n \to X$, thus the uniquely determined homomorphism

$$f_* : C_n(X; G) \to C_n(Y; G) : \sum_i a_i \sigma_i \mapsto \sum_i a_i (f \circ \sigma_i)$$

defines a chain map. □

---

[37] Since $\pi_1(X)/[\pi_1(X), \pi_1(X)]$ is abelian, we are adopting the convention of writing its group operation as addition, so the multiplication of an integer $m \in \mathbb{Z}$ by an element $\Psi(\sigma) \in \pi_1(X)/[\pi_1(X), \pi_1(X)]$ is defined accordingly.

Notice that the chain maps in the above lemma also satisfy $(f \circ g)_* = f_* \circ g_*$ whenever f and g are composable continuous maps, and the chain map induced by the identity map on X is simply the identity homomorphism on $C_*(X; G)$. Applying Proposition 22.5 thus gives the following result, which implies that homeomorphic spaces always have isomorphic singular homology groups:

Corollary 23.2. *Continuous maps* $f : X \to Y$ *determine group homomorphisms* $f_* : H_n(X; G) \to H_n(Y; G)$ *for every* n *and* G *such that* $(f \circ g)_* = f_* \circ g_*$ *whenever* f *and* g *can be composed, and the identity map satisfies* $(\mathrm{Id})_* = \mathbb{1}$. □

Remark 23.3. Recall that in the analogue of Corollary 23.2 for the fundamental group, the map $f : X \to Y$ is required to be base-point preserving, due to the fact that the definitions of $\pi_1(X)$ and $\pi_1(Y)$ require choices of base points in X and Y respectively. In most applications, base points are an extra piece of data that one doesn't actually care about but needs to keep track of anyway. One of the advantages of singular homology in comparison with the fundamental group is that its definition does not require any choice of base point, and Corollary 23.2 thus holds for *arbitrary* continuous maps $f : X \to Y$.

We will show in the next lecture that the homomorphisms f_* induced by continuous maps f only depend on f up to homotopy, which has the easy consequence that $H_*(X; G)$ only depends on the homotopy type of X.

But first, let us generalize the discussion somewhat. Algebraic gadgets often have the feature that they become easier to compute if you add more structure to them, sometimes at the cost of making the basic definitions slightly more elaborate. We will now do that with singular homology by introducing the *relative homology* groups of pairs. A **pair of spaces** (X, A), often abbreviated as simply a "pair," (*topologisches Paar*) consists of a topological space X and a subset $A \subset X$. Given two pairs (X, A) and (Y, B), a map $f : X \to Y$ is called a **map of pairs** if $f(A) \subset B$, and in this case we write

$$f : (X, A) \to (Y, B).$$

This is an obvious generalization of the definition of a pointed map, where arbitrary subsets have now replaced base points. Similarly, two maps of pairs $f, g : (X, A) \to (Y, B)$ are **homotopic** if there exists a homotopy $H : I \times X \to Y$ between f and g such that $H(s, \cdot) : (X, A) \to (Y, B)$ is a map of pairs for every $s \in I$, or equivalently,

$$H(I \times A) \subset B.$$

Two pairs (X, A) and (Y, B) are **homeomorphic** if there exist maps of pairs $f : (X, A) \to (Y, B)$ and $g : (Y, B) \to (X, A)$ such that $g \circ f$ and $f \circ g$ are the identity maps on (X, A) and (Y, B) respectively, and f and g are in this case called **homeomorphisms of pairs**. If $g \circ f$ and $f \circ g$ are not necessarily equal but are homotopic (as maps of pairs) to the respective identity maps, then we call each of them a **homotopy equivalence of pairs** and say that (X, A) and (Y, B) are homotopy equivalent, written

$$(X, A) \underset{h.e.}{\simeq} (Y, B).$$

One can regard every individual space X as a pair by identifying it with $(X, \varnothing)$, in which case the above definitions reproduce the usual ones for maps between ordinary spaces.

The relative homology of a pair (X, A) is based on the trivial observation that since every singular simplex in A is also a singular simplex in X whose boundary faces are all contained in A, $C_n(A; G)$ is naturally a subgroup of $C_n(X; G)$ for each n, and the boundary map $\partial : C_n(X; G) \to C_{n-1}(X; G)$ sends $C_n(A; G)$ to $C_{n-1}(A; G)$. It follows that ∂ descends to a sequence of well-defined homomorphisms on the quotients

$$C_n(X, A; G) := C_n(X; G)/C_n(A; G),$$

and since ∂^2 is still zero, $(C_*(X,A;G),\partial)$ is a chain complex, called the **relative singular chain complex** of the pair (X,A) with coefficients in G. Its homology groups are the **relative singular homology** (*relative singuläre Homologie*),

$$H_n(X,A;G) := H_n\left(C_*(X,A;G),\partial\right).$$

The case $A=\varnothing$ reproduces $H_n(X;G)$ as we defined it in the previous lecture, and these are sometimes called the **absolute** homology groups of X so as to distinguish them from relative homology groups. As in absolute homology, we may sometimes abbreviate the case of integer coefficients by

$$H_n(X,A) := H_n(X,A;\mathbb{Z}).$$

Lemma 23.1 extends in an obvious way to the relative chain complex: if $f:(X,A)\to(Y,B)$ is a map of pairs, then the absolute chain map $f_*:C_*(X;G)\to C_*(Y;G)$ sends the subgroup $C_*(A;G)$ into $C_*(B;G)$ and thus descends to a chain map

$$f_*:C_*(X,A;G)\to C_*(Y,B;G),$$

implying the relative version of Corollary 23.2:

Theorem 23.4. *Maps of pairs* $f:(X,A)\to(Y,B)$ *determine group homomorphisms* $f_*: H_n(X,A;G)\to H_n(Y,B;G)$ *for every* n *and* G *such that* $(f\circ g)_* = f_*\circ g_*$ *whenever* f *and* g *can be composed, and the identity map on* (X,A) *induces the identity homomorphism on* $H_n(X,A;G)$. ☐

Since $C_n(X,A;G)$ is a quotient, its elements are technically equivalence classes, but in order to avoid having too many equivalence relations floating around in the same discussion, let us instead think of them as ordinary n-chains $c\in C_n(X;G)$, keeping in mind that two such n-chains $a,b\in C_n(X;G)$ define the same element of $C_n(X,A;G)$ whenever $a-b\in C_n(A;G)$, meaning a and b differ by a linear combination of simplices that are all contained in A. A chain $c\in C_n(X;G)$ can then be called a **relative cycle** if the element of $C_n(X,A;G)$ it determines is a cycle, which means ∂c belongs to $C_{n-1}(A;G)$. Notice that a relative cycle need not be an **absolute cycle** in general (meaning $\partial c = 0$), though absolute cycles also define relative cycles. Relative cycles $c\in C_n(X;G)$ define relative homology classes $[c]\in C_n(X,A;G)$, and two relative cycles $b,c\in C_n(X;G)$ are homologous (meaning $[b]=[c]\in H_n(X,A;G)$) if and only if

$$b-c = a+\partial x \qquad \text{for some } a\in C_n(A;G),\ x\in C_{n+1}(X;G).$$

In particular, a relative cycle is nullhomologous if and only if it is the sum of a boundary plus a chain contained in A. If you find these algebraic relations overly abstract and would like some advice on how to actually *visualize* relative cycles, see the extended digression at the end of this lecture.

The reason for introducing the relative homology groups $H_*(X,A;G)$ was *not* that we wanted a tool for distinguishing non-homeomorphic pairs—the relative homology is such a tool, but our primary interest remains the space X on its own, rather than the pair (X,A). The usefulness of relative homology lies in the fact that there is a relation between the three groups $H_*(X;G)$, $H_*(A;G)$ and $H_*(X,A;G)$ for any pair (X,A), and indeed, one might hope to encounter situations in which two out of these three groups are easy to compute, so that a computation of the third one then comes for free. Let's make this idea more precise.

We begin with a seemingly trivial observation: let $i:A\hookrightarrow X$ and $j:X=(X,\varnothing)\hookrightarrow(X,A)$ denote the natural inclusions,[38] and consider the sequence of chain maps

$$0\longrightarrow C_*(A;G)\xrightarrow{i_*}C_*(X;G)\xrightarrow{j_*}C_*(X,A;G)\to 0, \tag{23.1}$$

---

[38] Strictly speaking, j in this context is just the identity map on X, but we cannot call it that since we are viewing it as a map between two non-identical pairs of spaces. It is a map of pairs due to the trivial fact that $\varnothing\subset A$.

where the first and last maps are each trivial. The map j_* is obviously surjective, as it is actually just the quotient projection

$$C_*(X;G) \to C_*(X,G)/C_*(A;G) = C_*(X,A;G).$$

The map i_* is similarly the inclusion $C_*(A;G) \hookrightarrow C_*(X;G)$ and is thus injective, and its image is precisely the kernel of j_*. This means that every term in this sequence has the property that the image of the preceding map equals the kernel of the next one. In general, a sequence of abelian groups with homomorphisms

$$\ldots \longrightarrow A_{n-2} \xrightarrow{f_{n-2}} A_{n-1} \xrightarrow{f_{n-1}} A_n \xrightarrow{f_n} A_{n+1} \xrightarrow{f_{n+1}} A_{n+2} \longrightarrow \ldots$$

is called **exact** (*exakt*) if $\ker f_n = \operatorname{im} f_{n-1}$ for every $n \in \mathbb{Z}$. If all the groups except for two neighboring groups in the sequence are trivial, then it suffices to look at a sequence of four groups with only one nontrivial homomorphism

$$0 \longrightarrow A_1 \xrightarrow{f} A_2 \longrightarrow 0,$$

and the exactness of the sequence then simply means that $f : A_1 \to A_2$ is both injective and surjective, i.e. it is an isomorphism. In this sense, one can think of an exact sequence as a generalization of the notion of an isomorphism between two abelian groups. The next simplest case is what is called a **short exact sequence** (*kurze exakte Sequenz*), in which all except three of the groups and two of the homomorphisms are trivial,

$$0 \longrightarrow A_1 \xrightarrow{f_1} A_2 \xrightarrow{f_2} A_3 \longrightarrow 0.$$

Exactness in this case means three things: f_1 is injective, f_2 is surjective, and $\operatorname{im} f_1 = \ker f_2$. The sequence in (23.1) is what we call a **short exact sequence of chain maps**, because the abelian groups in each term are also chain complexes and the homomorphisms between them are chain maps. One can now wonder what happens if we replace these chain complexes with their homology groups and the chain maps with the induced homomorphisms on homology: will the resulting sequence be exact? The answer is no, but what is actually true is much better and more useful than this:

Theorem 23.5. *Suppose (A_*, ∂^A), (B_*, ∂^B) and (C_*, ∂^C) are chain complexes and*

$$0 \longrightarrow A_* \xrightarrow{f} B_* \xrightarrow{g} C_* \longrightarrow 0$$

is a short exact sequence of chain maps. Then there exists a natural homomorphism $\partial_* : H_n(C_*, \partial^C) \to H_{n-1}(A_*, \partial^A)$ for each $n \in \mathbb{Z}$ such that the sequence

$$\begin{aligned} \ldots \xrightarrow{\partial_*} H_{n+1}(A_*, \partial^A) \xrightarrow{f_*} H_{n+1}(B_*, \partial^B) \xrightarrow{g_*} H_{n+1}(C_*, \partial^C) \\ \xrightarrow{\partial_*} H_n(A_*, \partial^A) \xrightarrow{f_*} H_n(B_*, \partial^B) \xrightarrow{g_*} H_n(C_*, \partial^C) \\ \xrightarrow{\partial_*} H_{n-1}(A_*, \partial^A) \xrightarrow{f_*} H_{n-1}(B_*, \partial^B) \xrightarrow{g_*} H_{n-1}(C_*, \partial^C) \xrightarrow{\partial_*} \ldots \end{aligned} \tag{23.2}$$

is exact.

The sequence of homology groups in this theorem is called a **long exact sequence** (*lange exakte Sequenz*), and the maps $\partial_* : H_n(C_*, \partial^C) \to H_{n-1}(A_*, \partial^A)$ are called the **connecting homomorphisms** in this sequence. In particular, this result turns (23.1) into the so-called **long exact sequence of the pair** (X, A),

$$\ldots \to H_{n+1}(X,A;G) \xrightarrow{\partial_*} H_n(A;G) \xrightarrow{i_*} H_n(X;G) \xrightarrow{j_*} H_n(X,A;G) \xrightarrow{\partial_*} H_{n-1}(A;G) \to \ldots \tag{23.3}$$

To see why this might be useful, notice what it implies if we happen to know for some reason that one of the three groups $H_n(X;G)$, $H_n(A;G)$ or $H_n(X,A;G)$ is trivial for every n; for concreteness, let's suppose it is known that $H_*(X, A; G) = 0$. This knowledge turns the long exact sequence (23.3) into an infinite collection of two-term exact sequences

$$0 \longrightarrow H_n(A; G) \xrightarrow{i_*} H_n(X; G) \longrightarrow 0,$$

implying that for every n, the map $i_* : H_n(A; G) \to H_n(X; G)$ is an isomorphism. If we are also lucky enough to know already what $H_*(A; G)$ is, then the computation of $H_*(X; G)$ is thus complete. An argument of this type will be used in Lecture 25 as the final step in computing $H_*(S^n; \mathbb{Z})$ for every $n \geqslant 1$.

Theorem 23.5 is a purely algebraic statement, and it is proved by a straightforward but nonetheless slightly surprising procedure known as "diagram chasing". I will not give the full argument here, because that would bore you to tears, but I will explain the first couple of steps, and I highly recommend that you work through the rest yourself the next time you are half-asleep and in need of amusement on an airplane, or recovering from surgery on heavy pain medication, as the case may be.[39] The basic idea is to write down a great big commutative diagram, examine at each step exactly what information you can deduce from exactness and commutativity, and then let the diagram tell you what to do.

Here is the diagram we need—it commutes because f and g are chain maps, and each of its rows is an exact sequence of abelian groups:

$$\begin{tikzcd}
 & \vdots \arrow[d, "\partial^A"] & \vdots \arrow[d, "\partial^B"] & \vdots \arrow[d, "\partial^C"] & \\
0 \arrow[r] & A_{n+1} \arrow[r, "f"] \arrow[d, "\partial^A"] & B_{n+1} \arrow[r, "g"] \arrow[d, "\partial^B"] & C_{n+1} \arrow[r] \arrow[d, "\partial^C"] & 0 \\
0 \arrow[r] & A_n \arrow[r, "f"] \arrow[d, "\partial^A"] & B_n \arrow[r, "g"] \arrow[d, "\partial^B"] & C_n \arrow[r] \arrow[d, "\partial^C"] & 0 \\
0 \arrow[r] & A_{n-1} \arrow[r, "f"] \arrow[d, "\partial^A"] & B_{n-1} \arrow[r, "g"] \arrow[d, "\partial^B"] & C_{n-1} \arrow[r] \arrow[d, "\partial^C"] & 0 \\
0 \arrow[r] & A_{n-2} \arrow[r, "f"] \arrow[d, "\partial^A"] & B_{n-2} \arrow[r, "g"] \arrow[d, "\partial^B"] & C_{n-2} \arrow[r] \arrow[d, "\partial^C"] & 0 \\
 & \vdots & \vdots & \vdots &
\end{tikzcd}$$

We start by writing down a reasonable candidate for the map $\partial_* : H_n(C_*, \partial^C) \to H_{n-1}(A_*, \partial^A)$. Given $[c] \in H_n(C_*, \partial^C)$, $c \in C_n$ is necessarily a cycle, and exactness tells us that $g : B_n \to C_n$ is surjective, hence $c = g(b)$ for some $b \in B_n$. Then using commutativity,

$$0 = \partial^C c = \partial^C g(b) = g(\partial^B b),$$

so $\partial^B b \in \ker g \subset B_{n-1}$, and using exactness again, this implies $\partial^B b = f(a)$ for some $a \in A_{n-1}$. Notice that a is uniquely determined by b since (using exactness again) f is injective. Applying commutativity again, we now observe that

$$f(\partial^A a) = \partial^B(f(a)) = \partial^B \partial^B b = 0$$

---

[39] I first learned about exact sequences around the same time that I had all four of my wisdom teeth removed in a complicated procedure that left me drowsily dependent on prescription pain medication for about three weeks afterward. It turns out that that was exactly the right frame of mind in which to work through diagram chasing arguments without getting bored.

since $(\partial^B)^2 = 0$, and the injectivity of f then implies $\partial^A a = 0$. So just by chasing the diagram from C_n to A_{n-1}, we found a cycle $a \in A_{n-1}$, and it seems reasonable to define

$$\partial_*[c] := [a] \in H_{n-1}(A, \partial^A).$$

We need to check that this is well defined, as two arbitrary choices were made in the procedure going from $[c]$ to $[a]$. One was the choice of an element $b \in B_n$ with $g(b) = c$, so we could get a different cycle $a' \in A_{n-1}$ by choosing a different element $b' \in g^{-1}(c)$ and requiring $f(a') = \partial^B b'$. But then $b' - b$ belongs to $\ker g = \operatorname{im} f$, hence we can write $b' - b = f(x)$ for some $x \in A_n$, implying

$$f(a' - a) = f(a') - f(a) = \partial^B(b' - b) = \partial^B(f(x)) = f(\partial^A(x)),$$

and since f is injective, $a' - a = \partial^A x$, implying that a and a' are homologous cycles. The other choice we made was the cycle $c \in C_n$, which in principle we are free to replace by any homologous cycle $c' \in C_n$ and then follow the same procedure to produce a different cycle $a' \in A_{n-1}$. If we do this, then $c' - c = \partial^C z$ for some $z \in C_{n+1}$, and since g is surjective, $z = g(y)$ for some $y \in B_{n+1}$. We then have

$$c' - c = \partial^C(g(y)) = g(\partial^B(y)),$$

and since we now know that we are free to choose any $b \in g^{-1}(c)$ and $b' \in g^{-1}(c')$, we can set

$$b' := b + \partial^B(y).$$

This implies $\partial^B b' = \partial^B b$, thus the condition $f(a') = \partial^B b'$ produces $a' = a$, and we have finished the proof that ∂_* is well defined.

It remains to prove that ∂_* really is a homomorphism, and that the long exact sequence really is exact, i.e. that $\ker \partial_* = \operatorname{im} g_*$, $\ker g_* = \operatorname{im} f_*$ and $\ker f_* = \operatorname{im} \partial_*$. This can all be done by the same kinds of straightforward arguments as above, but I'm sure you can see now why I'm not going to write down the complete details here.

I have one final remark however about the long exact sequence of a pair (X, A). If you redo the diagram chase above for the particular short exact sequence (23.1), you end up with a precise and very natural formula for the connecting homomorphisms

$$\partial_* : H_n(X, A; G) \to H_{n-1}(A; G).$$

The procedure starts with a relative n-cycle $c \in C_n(X, A; G)$, from which we need to pick $b \in j_*^{-1}(c) \subset C_n(X; G)$, but if we apply the usual convention of regarding relative cycles in (X, A) as chains in X, then c is already in $C_n(X; G)$ and we can pick b to be exactly the same chain c. Next we look at $\partial c \in C_{n-1}(X; G)$ and find the unique cycle $a \in C_{n-1}(A; G)$ that is sent to ∂c under the inclusion $C_{n-1}(A; G) \hookrightarrow C_{n-1}(X; G)$. In other words, $a = \partial c$, so the "obvious" formula is the right one:

$$\partial_*[c] = [\partial c]. \tag{23.4}$$

This looks more trivial than it is, e.g. you might think that $[\partial c]$ should automatically be 0 because ∂c is a boundary, but the point is that c is a chain in X, it might not be confined to A, so ∂c is certainly a cycle in A (as a consequence of the fact that c is a relative chain in (X, A)) but it need not be the boundary of any chain in A, and $[\partial c]$ may very well be a nontrivial homology class in $H_{n-1}(A; G)$.

Exercise 23.6. Use the formula (23.4) to give a direct proof that the sequence (23.3) is exact.

Remark 23.7. Exercise 23.6 is straightforward and doable in a much shorter time than the proof of Theorem 23.5, so we could have skipped the abstract homological algebra discussion without losing anything that is essential for the current semester. However, I wanted to make the point that the long exact sequence of a pair is not just an isolated topological phenomenon—it is a special case of a much more general algebraic principle, and that principle reappears in many other contexts in various branches of mathematics. We will see it again several times in *Topologie II*.

The following **extended digression** is not logically necessary for our development of basic homology theory, but you might still appreciate some intuition on the following question: what do relative n-cycles actually *look like*? Actually, that's also a valid question when applied to absolute n-cycles, and we've only really addressed it so far for $n = 0$ and $n = 1$. The best way I know for visualizing absolute cycles is via the analogy with bordism theory. Recall that elements of $\Omega_n^{\mathrm{SO}}(X)$ are equivalence classes of maps $f : M \to X$ where M is a closed oriented n-manifold. If M admits an oriented triangulation, then after choosing an ordering for all the vertices in this triangulation and assigning orientations accordingly to each simplex in the triangulation, one can identify each k-simplex $\sigma \subset M$ with a map $\Delta^k \to M$ that parametrizes it, thus defining a singular k-simplex in M. For $k = n$ in particular, the condition in Definition 20.9 relating the orientations of neighboring n-simplices implies that the sum $\sum_i \epsilon_i \sigma_i$ of all the singular n-simplices in the triangulation—with appropriate signs $\epsilon_i = \pm 1$ attached in order to describe their orientations in the triangulation—is a cycle in $C_n(M; \mathbb{Z})$. This is true because in $\partial \sum_i \epsilon_i \sigma_i$, every $(n-1)$-simplex of the triangulation appears exactly twice, but the orientation condition requires these two instances to appear with opposite signs. The resulting singular homology class is denoted by

$$[M] := \left[\sum_i \epsilon_i \sigma_i \right] \in H_n(M; \mathbb{Z})$$

and called the **fundamental class** (*Fundamentalklasse*) of M. We cannot prove it right now, but we will see in *Topologie II* that $[M]$ does not depend on the choice of triangulation, and it can even be defined for arbitrary closed and oriented topological manifolds, which need not admit triangulations. The map $f : M \to X$ then determines a corresponding cycle $\sum_i \epsilon_i (f \circ \sigma_i) \in C_n(X; \mathbb{Z})$ and an n-dimensional homology class $f_*[M] \in H_n(X; \mathbb{Z})$.

How can we recognize when two n-cycles in X defined in this way are homologous, or equivalently, when $\sum_i \epsilon_i (f \circ \sigma_i)$ is nullhomologous? A nice answer can again be extracted from bordism theory. If $[(M, f)] = 0 \in \Omega_n^{\mathrm{SO}}(X)$, it means there exists a compact oriented $(n+1)$-manifold W with $\partial W \cong M$ and a map $F : W \to X$ with $F|_M = f$. Suppose W admits an oriented triangulation that restricts to ∂W as an oriented triangulation of M. Identifying the $(n+1)$-simplices τ_j in this triangulation with singular $(n+1)$-simplices in W and then adding them up with suitable signs $\epsilon_j = \pm 1$ as in the previous paragraph produces an $(n+1)$-chain in X of the form $\sum_j \epsilon_j (F \circ \tau_j)$, whose boundary is the n-cycle representing $f_*[M]$. Thus if oriented triangulations can always be assumed to exist, then $f_*[M] = 0 \in H_n(X; \mathbb{Z})$ whenever (M, f) is nullbordant, and similarly, $f_*[M] = g_*[N] \in H_n(X; \mathbb{Z})$ will hold whenever (M, f) and (N, g) are related by an oriented bordism. We will also see in *Topologie II* that these statements remain true without mentioning triangulations.

You may be wondering how general this discussion really is, i.e. does *every* integral homology class in X arise from a map of a closed manifold into X? The answer is in general no, but if X is a nice enough space like the polyhedron of a finite simplicial complex, then something almost as good is true. The proof of the following famous result of Thom would be far beyond the scope of this course, and we will not make use of it, but it is nice to know that it exists.

THEOREM 23.8 (R. Thom [Tho54]). *If X is a compact polyhedron, then for every $n \geqslant 0$ and $A \in H_n(X; \mathbb{Z})$, there exists a closed n-manifold M, a map $f : M \to X$ and a number $k \in \mathbb{N}$ such that $kA = f_*[M]$.* ☐

To talk about relative homology classes, we could now allow M to be a compact oriented n-manifold with boundary and assume that its oriented triangulation also defines an oriented triangulation of ∂M. The chain $\sum_i \epsilon_i \sigma_i \in C_n(M;\mathbb{Z})$ is then no longer a cycle, because $(n-1)$-simplices on ∂M are not canceled, they each appear exactly once. Instead, $\partial \sum_i \epsilon_i \sigma_i$ is an $(n-1)$-cycle representing the fundamental class of ∂M, and $\sum_i \epsilon_i \sigma_i$ is therefore a relative cycle in $(M, \partial M)$, defining a **relative fundmental class**

$$[M] \in H_n(M, \partial M; \mathbb{Z}).$$

Given a pair (X, A), any map $f : (M, \partial M) \to (X, A)$ now determines a relative cycle $\sum_i \epsilon_i (f \circ \sigma_i) \in C_n(X, A; \mathbb{Z})$ and relative homology class $f_*[M] \in H_n(X, A; \mathbb{Z})$. For intuition, it is usually helpful to assume that f is an embedding, so a relative n-cycle in (X, A) then looks like an oriented and triangulated compact n-dimensional submanifold in X whose boundary lies in A.

Finally, note that one can drop the orientations from this entire discussion at the cost of replacing $\mathbb{Z}$ coefficients with $\mathbb{Z}_2$. Indeed, if M is closed and has a triangulation but not one that is orientable, then the n-chain defined by adding up the n-simplices may not be a cycle because its boundary may include some $(n-1)$-simplex that appears twice without canceling. But since $2 = 0 \in \mathbb{Z}_2$, this sum still defines a cycle in $C_n(M; \mathbb{Z}_2)$ and therefore also a fundamental class

$$[M] \in H_n(M; \mathbb{Z}_2).$$

This reveals that unoriented bordism classes in $\Omega_n(X)$ determine homology classes in $H_n(X; \mathbb{Z}_2)$, and the analogue of Theorem 23.8 remains true in this case without any need for the multiplicative factor $k \in \mathbb{N}$.

24. Homotopy invariance and excision

We need to prove two more theorems about singular homology before it becomes a truly useful tool. Both will require a bit of work, but the almost immediate payoff will be that we can then compute the homology of spheres in every dimension. This has several important applications, including the general case of the Brouwer fixed point theorem, and the basic fact that open sets in $\mathbb{R}^n$ are never homeomorphic to open sets in $\mathbb{R}^m$ unless $n = m$. It is also the first step in developing an algorithm to compute the singular homology of any CW-complex, a general class of "reasonable" spaces that includes all smooth manifolds and all simplicial complexes.

Our first task for today is homotopy invariance.

THEOREM 24.1. *The map $f_* : H_n(X, A; G) \to H_n(Y, B; G)$ induced for each $n \in \mathbb{Z}$ by a map of pairs $f : (X, A) \to (Y, B)$ depends only on the homotopy class of f (as a map of pairs).*

The obvious corollary about homotopy equivalent spaces is a result of tremendous theoretical importance, and I would like to point out how much simpler its proof is than that of the corresponding statement about fundamental groups (Theorem 10.23). The complication in the case of π_1 was that its definition depends on a choice of base point, but the notion of homotopy equivalence does not—as a result, we had to find a workaround to cope with the fact that homotopy inverses need not be base-point preserving. In homology, one can also allow for base points by considering pairs (X, A) where $A \subset X$ is a single point, but homotopies between maps of pairs are required to respect this extra data, which makes the proofs easier. And unlike the fundamental group, homology also makes sense for pairs (X, A) with $A = \varnothing$, in which case the terms "homotopy" and "homotopy equivalence" mean the same thing that they always did.

COROLLARY 24.2. *If $f : (X, A) \to (Y, B)$ is a homotopy equivalence of pairs, then the induced maps $f_* : H_n(X, A; G) \to H_n(Y, B; G)$ are isomorphisms.*

PROOF. Suppose $f : (X, A) \to (Y, B)$ is a homotopy equivalence, so it has a homotopy inverse $g : (Y, B) \to (X, A)$. Then $f \circ g$ and $g \circ f$ are homotopic to the identity maps on (Y, B) and (X, A)

respectively, so that Theorem 24.1 gives $f_* \circ g_* = \mathbb{1}$ and $g_* \circ f_* = \mathbb{1}$ for the induced maps on homology, implying that both are isomorphisms. □

The proof of Theorem 24.1 requires another fundamental notion from homological algebra. It should be clear that if $f, g : X \to Y$ are two non-identical maps, then the induced chain maps $f_*, g_* : C_*(X; G) \to C_*(Y; G)$ will not be identical, even if f and g are homotopic. It is still possible however for two distinct chain maps to descend to exactly the same map between homology groups. What we need for Theorem 24.1 is an algebraic mechanism to recognize when this happens, and that mechanism is called *chain homotopy.*

Definition 24.3. A **chain homotopy** (*Kettenhomotopie*) between two chain maps $f, g : (A_*, \partial^A) \to (B_*, \partial^B)$ is a sequence of homomorphisms $h_n : A_n \to B_{n+1}$ such that for every $n \in \mathbb{Z}$,

$$f_n - g_n = \partial^B_{n+1} \circ h_n + h_{n-1} \circ \partial^A_n.$$

In other words, a chain homotopy between f and g is a homomorphism $h : A_* \to B_*$ of degree $+1$ such that $f - g = \partial^B \circ h + h \circ \partial^A$. We sometimes abuse notation and write

$$h : A_* \to B_{*+1}$$

to emphasize that a chain homotopy is a homomorphism of degree 1.

Two chain maps that admit a chain homotopy between them are called **chain homotopic** (*kettenhomotop*), and it is not hard to show that this defines an equivalence relation on chain maps. You can picture a chain homotopy as a sequence of down-left diagonal arrows in the diagram (22.1), though you need to be a little careful with that diagram since a chain homotopy does not make it commute. The main importance of chain homotopies comes from the following result.

Proposition 24.4. *If there exists a chain homotopy between two chain maps f and g from (A_*, ∂^A) to (B_*, ∂^B), then they induce the same homomorphisms*

$$f_* = g_* : H_n(A_*, \partial^A) \to H_n(B_*, \partial^B)$$

for all $n \in \mathbb{Z}$.

Proof. If $h : A_* \to B_{*+1}$ is a chain homotopy, then given any $[a] \in H_n(A_*, \partial^A)$, we have $\partial^A a = 0$ and thus

$$f(a) - g(a) = \partial^B h(a) + h(\partial^A a) = \partial^B (h(a)),$$

hence $f(a)$ and $g(a)$ are homologous cycles. □

If you're seeing the notion of chain homotopies for the first time, you might think that the definition above looks a bit unmotivated—it is not obvious for instance whether this is the *only* reasonable algebraic condition that makes two chain maps induce the same map on homology. However, the following lemma and its proof provide convincing evidence that this definition is the right one: it turns out that chain homotopies are the *natural* algebraic structure that arises in the singular chain complex from a homotopy between continuous maps. We will see that they arise naturally in many other contexts as well.

Lemma 24.5. *If there exists a homotopy between the maps of pairs $f, g : (X, A) \to (Y, B)$, then there also exists a chain homotopy between the induced chain maps $f_*, g_* : C_*(X, A; G) \to C_*(Y, B; G)$.*

Theorem 24.1 is an immediate consequence of this lemma and Proposition 24.4, so our remaining task is to prove the lemma. For notational simplicity, let us start under the assumption

$$A = B = \varnothing,$$

as the general case will only require a few extra remarks beyond this. Suppose $H : I \times X \to Y$ is a homotopy between $f = H(0, \cdot)$ and $g = H(1, \cdot)$. Associate to each singular n-simplex $\sigma : \Delta^n \to X$ the map

$$h_\sigma : I \times \Delta^n \to Y : (s, t) \mapsto H(s, \sigma(t)),$$

so $h_\sigma(0, \cdot) = f \circ \sigma$ and $h_\sigma(1, \cdot) = g \circ \sigma$. If we pretend for a moment that the maps in this picture are all embeddings, then we can picture h_σ as tracing out a "prism-shaped" region in Y whose boundary consists of three pieces, two of which are the n-simplices traced about by $f_*\sigma$ and $g_*\sigma$. If we pay proper attention to orientations, then $f_*\sigma$ will get a negative orientation because the boundary orientation for $\partial(I \times \Delta^n)$ induces opposite orientations on $\{0\} \times \Delta^n$ and $\{1\} \times \Delta^n$. But there is a third piece of $\partial(I \times \Delta^n)$ that we haven't mentioned yet, namely $I \times \partial\Delta^n$. If we regard $I \times \Delta^n$ as a compact oriented $(n + 1)$-manifold with boundary, then its oriented boundary turns out to be[40]

$$\partial(I \times \Delta^n) = (-\{0\} \times \Delta^n) \cup (\{1\} \times \Delta^n) \cup (-I \times \partial\Delta^n) . \tag{24.1}$$

This relation will be the geometric motivation behind the chain homotopy formula.

The idea now is to define a chain homotopy $h : C_*(X; G) \to C_{*+1}(Y; G)$ by associating to each singular n-simplex $\sigma : \Delta^n \to X$ a linear combination of singular $(n+1)$-simplices in Y determined by the prism map $h_\sigma : I \times \Delta^n \to Y$. Unfortunately, $I \times \Delta^n$ is not a simplex, but there are various natural ways to decompose it into simplices, i.e. to triangulate it. In principle, the result should not depend on how this is done, so long as the triangulation has reasonable properties, thus we will not explain the details here except to state what properties are needed:

LEMMA 24.6. *There exists a sequence of oriented triangulations of the sequence of spaces* $I \times \Delta^n$ *for* $n = 0, 1, 2, \dots$ *satisfying the following properties:*

1. $\{0\} \times \Delta^n$ *and* $\{1\} \times \Delta^n$ *are boundary faces of* $(n + 1)$*-simplices in the triangulation of* $I \times \Delta^n$*;*
2. *Under the natural identification of each boundary face* $\partial_{(k)}\Delta^n$ *with* Δ^{n-1}*, the triangulation of* $I \times \Delta^n$ *restricts to* $I \times \partial_{(k)}\Delta^n$ *as the triangulation of* $I \times \Delta^{n-1}$*.*

A precise algorithm to produce such triangulations of $I \times \Delta^n$ is described in [**Hat02**, p. 112]. I recommend taking a moment to draw pictures of how it might be done for $n = 1$ and $n = 2$. In the following, we will assume that parametrizations $\tau_i : \Delta^{n+1} \to I \times \Delta^n$ of the finite set of $(n + 1)$-simplices in these triangulations have also been chosen such that for a suitable choice of signs $\epsilon_i = \pm 1$ determined by their orientations,

$$\sum_i \epsilon_i \tau_i \in C_{n+1}(I \times \Delta^n; \mathbb{Z})$$

defines a relative cycle in $(I \times \Delta^n, \partial(I \times \Delta^n))$; in other words, all interior n-simplices in the triangulation of $I \times \Delta^n$ appear twice with opposite signs in $\partial \sum_i \epsilon_i\tau_i$, so that what remains is an n-chain in the boundary. The stated conditions on the triangulation guarantee in fact that $\partial \sum_i \epsilon_i\tau_i$ will consist of the following terms:

1. A single term for the obvious parametrization $\Delta^n \to \{1\} \times \Delta^n$, whose attached coefficient we can assume without loss of generality is $+1$;
2. Another term for the obvious parametrization $\Delta^n \to \{0\} \times \Delta^n$, whose attached coefficient must now be -1 for orientation reasons;

---

[40] One can deduce the signs in (24.1) from things that were said in Lecture 20, though it's a bit tedious, and for now I would encourage you to just believe me that the signs are correct. There is an easier way to see it using the notion of orientation for *smooth* manifolds and their tangent spaces, which we do not have space to talk about here, but you'll likely see things like this again in differential geometry at some point.

(3) Linear combinations (with coefficients ± 1) of the n-simplices triangulating $I \times \partial_{(k)}\Delta^n = I \times \Delta^{n-1}$ for each boundary face of Δ^n.

With this in hand, there is a unique homomorphism $h : C_n(X; G) \to C_{n+1}(Y; G)$ defined on each singular n-simplex $\sigma : \Delta^n \to X$ by the formula

$$h(\sigma) := \sum_i \epsilon_i (h_\sigma \circ \tau_i) \in C_{n+1}(Y; \mathbb{Z}),$$

where the sum is over all the parametrized $(n+1)$-simplices $\tau_i : \Delta^{n+1} \to I \times \Delta^n$ in our triangulation from Lemma 24.6, and the $\epsilon_i = \pm 1$ are determined by their orientations as outlined above. In light of (24.1), we then have

$$\partial h(\sigma) = g_* \sigma - f_* \sigma - h(\partial \sigma),$$

where the third term comes from the restriction of h_σ to the triangulated subset $-I \times \partial \Delta^n$ in the oriented boundary of $I \times \Delta^n$. It follows that $h : C_*(X; G) \to C_{*+1}(Y; G)$ satisfies $\partial \circ h + h \circ \partial = g_* - f_*$, i.e. h is a chain homotopy.

This concludes the proof of Lemma 24.5 in the case $A = B = \varnothing$. In the general case, the given homotopy satisfies the additional assumption

$$H(I \times A) \subset B,$$

thus following through with the above construction, h_σ has image contained in B whenever σ has image in A. It follows that the chain homotopy we constructed sends $C_n(A; G)$ into $C_{n+1}(B; G)$ and thus descends to the quotients as a chain homotopy

$$h_* : C_*(X, A; G) \to C_{*+1}(Y, B; G)$$

between the relative chain maps $f_*, g_* : C_*(X, A; G) \to C_*(Y, B; G)$. The proof of the lemma is now complete, and with it, the proof of the homotopy invariance of singular homology.

Let us pick some low-hanging fruit from this result.

Corollary 24.7 (via Exercise 22.9). *For any contractible space X and any coefficient group G, $H_n(X; G)$ is isomorphic to G for $n = 0$ and vanishes for $n \neq 0$.* ☐

Corollary 24.8 (via Theorem 22.10). *If X is homotopy equivalent to S^1, then $H_1(X; \mathbb{Z}) \cong \mathbb{Z}$.* ☐

The second big theorem for today is called the *excision* property. It is based on the intuition that since $H_*(X, A; G)$ is supposed to ignore anything that happens entirely inside the subset A, removing smaller subsets $B \subset A$ should not change the relative homology, i.e. we expect

$$H_*(X \backslash B, A \backslash B; G) \cong H_*(X, A; G).$$

This works under a mild assumption on what it means for a subset B to be "smaller" than A.

Theorem 24.9 (excision). *For any pair (X, A), if $B \subset A$ is a subset with closure contained in the interior of A, then the inclusion of pairs $i : (X \backslash B, A \backslash B) \hookrightarrow (X, A)$ induces isomorphisms*

$$i_* : H_n(X \backslash B, A \backslash B; G) \xrightarrow{\cong} H_n(X, A; G)$$

for all n and G.

The assumption $B \subset \bar{B} \subset \mathring{A} \subset A \subset X$ means essentially that the two open subsets $\mathring{A}$ and $X \backslash \bar{B}$ cover X. In this setting, let us say that a chain $c \in C_n(X; G)$ is *decomposable* if c can be written as a sum of a chain in A plus a chain in $X \backslash B$, i.e. c belongs to the subgroup $C_n(A; G) + C_n(X \backslash B; G) \subset C_n(X; G)$. The excision theorem is closely related to the observation that every relative n-cycle in (X, A) is homologous to one that is decomposable. Indeed, if this is true and every $[c] \in H_n(X, A; G)$ can be written without loss of generality as $c = c_A + c_{X \backslash B}$ for some $c_A \in C_n(A;G)$ and $C_{X\backslash B} \in C_n(X\backslash B;G)$, then since c is a relative cycle, $\partial c \in C_{n-1}(A;G)$, implying $\partial c_{X\backslash B}$ is also in $C_{n-1}(A;G)$ since ∂c_A must be as well, thus $\partial c_{X\backslash B} \in C_{n-1}(A\backslash B;G)$. This proves that $c_{X\backslash B}$ is a relative n-cycle for the pair $(X\backslash B, A\backslash B)$, so it represents a homology class in $H_n(X\backslash B, A\backslash B;G)$, and obviously

$$i_*[c_{X\backslash B}] = [c]$$

since $c_A \in C_n(A;G)$ represents the trivial element of $C_n(X,A;G)$. This proves surjectivity in Theorem 24.9, modulo the detail about why we are allowed to restrict our attention to decomposable chains. The latter is where most of the hard work is hidden.

Let us reframe the discussion slightly and suppose $\mathcal{U},\mathcal{V} \subset X$ are two subsets whose interiors form an open cover of X,

$$X = \mathring{\mathcal{U}} \cup \mathring{\mathcal{V}}.$$

We would like to develop a procedure for replacing any given chain $c \in C_n(X;G)$ with one that is in the subgroup $C_n(\mathcal{U};G) + C_n(\mathcal{V};G) \subset C_n(X;G)$ but represents the same homology class in cases where c is a (relative) cycle. If you followed the extended digression on how to visualize n-cycles at the end of the previous lecture, then you can imagine an intuitive reason why this should be possible: consider a homology class that is presented in the form $f_*[M] \in H_n(X;\mathbb{Z})$ for some triangulated oriented n-manifold M and a map $f : M \to X$. In this case, the definition of a cycle representing $f_*[M]$ depends on a choice of oriented triangulation for M, but we do not really expect the homology class $f_*[M]$ to depend on this triangulation, and in particular, we should be free to replace the triangulation by a *finer* one, which has more simplices but each one small enough to be contained in either $\mathcal{U}$ or $\mathcal{V}$ (or both). It is not hard to imagine that one could achieve this simply by triangulating each individual simplex in M to decompose it into strictly smaller simplices, and the process could then be repeated finitely many times to make the simplices as small as we like. This process is called *subdivision*. We shall now describe an inductive algorithm that makes the idea precise.

The **barycentric subdivision** of the standard n-simplex Δ^n is an oriented triangulation of Δ^n defined as follows. If $n = 0$, then Δ^0 is only a single point, so it cannot be subdivided any further and our triangulation of Δ^0 will consist only of that single 0-simplex. Now by induction, assume the desired triangulation of Δ^m has already been defined for all $m \leqslant n-1$. Under the natural identification of each boundary face $\partial_{(k)}\Delta^n$ with Δ^{n-1}, this means in particular that a triangulation of $\partial_{(k)}\Delta^n$ has been chosen for each $k = 0,\dots,n$. Now for each $(n-1)$-simplex σ in that triangulation, define σ' to be the n-simplex in Δ^n that is linearly spanned by the n vertices of σ plus one extra vertex that is in the interior of Δ^n, the so-called **barycenter**

$$b_n := \left(\frac{1}{n+1},\dots,\frac{1}{n+1}\right) \in \Delta^n.$$

It is straightforward to check that the collection of all n-simplices σ' defined in this way from $(n-1)$-simplices σ in boundary faces $\partial_{(k)}\Delta^n$ forms a triangulation of Δ^n, and one can also assign it an orientation based on the orientations of the triangulations of $\partial_{(k)}\Delta^n$. Some pictures for $n = 1,2,3$ are shown in [**Hat02**, p. 120].

As usual with triangulations of manifolds, one can assign to each n-simplex $\sigma' \subset \Delta^n$ in the barycentric subdivision of Δ^n a parametrization $\tau : \Delta^n \xrightarrow{\cong} \sigma' \subset \Delta^n$ such that the sum over all such parametrized simplices τ_i with attached signs $\epsilon_i = \pm 1$ determined by their orientations in the triangulation produces a relative n-cycle in $(\Delta^n, \partial\Delta^n)$,

$$\sum_i \epsilon_i\tau_i \in C_n(\Delta^n;\mathbb{Z}), \qquad \partial\sum_i \epsilon_i\tau_i \in C_{n-1}(\partial\Delta^n;\mathbb{Z}),$$

where $(n-1)$-simplices in the interior of Δ^n do not appear in $\partial\sum_i \epsilon_i\tau_i$ because each is a boundary face of two n-simplices whose induced boundary orientations cancel. We can then use this to define a homomorphism

$$S : C_n(X;G) \to C_n(X;G)$$

via the formula

$$S(\sigma) := \sum_i \epsilon_i(\sigma \circ \tau_i)$$

for each $n \geqslant 0$ and $\sigma : \Delta^n \to X$. Essentially, S replaces each singular n-simplex σ by a linear combination (with coefficients ± 1) of the restrictions of σ to the subdivided pieces of its domain.

Lemma 24.10. *$S : C_*(X;G) \to C_*(X;G)$ is a chain map.*

Proof. This follows from the relation $\partial S(\sigma) = S(\partial\sigma)$ for each $\sigma : \Delta^n \to X$, which is a direct consequence of the inductive nature of the subdivision algorithm: boundary faces of the smaller simplices in the subdivision are also the simplices in a subdivision of the original boundary faces. □

Lemma 24.11. *$S : C_*(X;G) \to C_*(X;G)$ is chain homotopic to the identity map.*

Proof. As in the proof of Lemma 24.5, the chain homotopy here comes from a particular choice of oriented triangulation of the prism $I \times \Delta^n$. A picture of this triangulation and a precise algorithm to construct it are given in [**Hat02**, p. 122]. We want it in particular to have the following properties:

(1) Its restriction to $\{1\} \times \Delta^n$ is the barycentric subdivision of Δ^n;
(2) Its restriction to $\{0\} \times \Delta^n$ consists only of that one n-simplex, with no subdivision;
(3) Its restriction to each $I \times \partial_{(k)}\Delta^n$ matches the chosen triangulation of $I \times \Delta^{n-1}$.

The third property means that the construction is again inductive: we start with $n = 0$ by choosing the trivial triangulation of $I \times \Delta^0 = I$, and then increase the dimension one at a time such that the triangulation already defined for $I \times \Delta^{n-1}$ determines the triangulation of $I \times \Delta^n$. Since it is an oriented triangulation, one can now define a relative $(n+1)$-cycle in $(I \times \Delta^n, \partial(I \times \Delta^n))$ of the form

$$\sum_i \epsilon_i \tau_i \in C_{n+1}(I \times \Delta^n; \mathbb{Z}),$$

where $\tau_i : \Delta^{n+1} \to I \times \Delta^n$ are parametrizations of the simplices in the triangulation and the signs $\epsilon_i = \pm 1$ are determined by their orientations. Let

$$\pi : I \times \Delta^n \to \Delta^n$$

denote the obvious projection map. The desired chain homotopy $h : C_n(X;G) \to C_{n+1}(X;G)$ is then determined by the formula

$$h(\sigma) = \sum_i \epsilon_i \left(\sigma \circ \pi \circ \tau_i\right).$$

In computing $\partial h(\sigma)$, n-simplices in the interior of $I \times \Delta^n$ make no contribution due to the usual cancelations, but there are contributions from the induced triangulation of $\partial(I \times \Delta^n)$, and the chain homotopy relation again follows from the geometric formula (24.1) for the oriented boundary of $I \times \Delta^n$. Namely, restricting to $\{1\} \times \Delta^n$ gives the barycentric subdivision $S(\sigma)$, restricting to $-\{0\} \times \Delta^n$ gives $-\sigma$, and restricting to $-I \times \partial\Delta^n$ gives the same operator applied to $\partial\sigma$, hence

$$\partial h(\sigma) = S(\sigma) - \sigma - h(\partial\sigma),$$

proving $S - \mathbb{1} = \partial h + h\partial$. □

The chain homotopy result implies that our subdivision map $S : C_*(X;G) \to C_*(X;G)$ has the main property we want, namely it induces the identity homomorphism $H_*(X;G) \to H_*(X;G)$, and since S clearly also preserves $C_*(A;G)$ for any $A \subset X$, the same is also true for the relative homology groups of (X,A). It then remains true if we replace S by any iteration S^m for integers $m \geqslant 1$, thus we can apply S repeatedly in order to make the individual simplices in a chain as small as we like. In particular, for any $c \in C_*(X;G)$, we will have $S^m c \in C_*(\mathcal{U};G) + C_*(\mathcal{V};G)$ for m sufficiently large. This is enough information to prove the excision theorem, so let's go ahead and do that.

PROOF OF THEOREM 24.9. The hypotheses of the theorem imply that X is the union of the interiors of $X\backslash B$ and A, so given any class $[c] \in H_n(X,A;G)$ with a relative n-cycle $c \in C_n(X;G)$ representing it, c can be replaced by an iterated subdivision $S^m c$ for large $m \in \mathbb{N}$ that represents the same relative homology class $[S^m c] = [c] \in H_n(X,A;G)$ but is also decomposable, meaning it is the sum of a chain in $X\backslash B$ with a chain in A. Let's assume that c has already been replaced with $S^m c$ in this way, so that without loss of generality,

$$c = c_A + c_{X\backslash B} \qquad \text{for some} \qquad c_A \in C_n(A;G),\; c_{X\backslash B} \in C_n(X\backslash B;G).$$

Having made this assumption, the reason why $i_* : H_n(X\backslash B, A\backslash B;G) \to H_n(X,A;G)$ is surjective was explained already in the paragraph after the statement of the theorem: the fact that $c \in C_n(X,A;G)$ is a relative n-cycle means $\partial c \in C_n(A;G)$ and therefore also $\partial c_{X\backslash B} \in C_n(A;G)$, so that $c_{X\backslash B}$ is a relative n-cycle in $(X\backslash B, A\backslash B)$, thus representing a class $[c_{X\backslash B}] \in H_n(X\backslash B, A\backslash B;G)$ that satisfies

$$i_*[c_{X\backslash B}] = [c].$$

The proof that $i_* : H_n(X\backslash B, A\backslash B;G) \to H_n(X,A;G)$ is injective uses subdivision in a slightly different way. Suppose $c \in C_n(X\backslash B;G)$ is a relative n-cycle representing a homology class $[c] \in H_n(X\backslash B, A\backslash B;G)$ with $i_*[c] = 0 \in H_n(X,A;G)$. Since i is just an inclusion map, $i_*[c] = 0$ means that after reinterpreting c as an n-chain in X instead of just in $X\backslash B$, c is a boundary of some $(n+1)$-chain in X, modulo one that is contained in A, i.e. we have

$$c = \partial b + a \qquad \text{for some } b \in C_{n+1}(X;G) \text{ and } a \in C_n(A;G).$$

Applying ∂ to both sides of this equation gives $\partial c = \partial a$, which implies since c is a relative n-cycle in $(X\backslash B, A\backslash B)$ that $\partial a \in C_n(A\backslash B;G)$, i.e. none of the singular simplices that make up the $(n-1)$-cycle ∂a intersect B. If we happened to know that the chains $b \in C_{n+1}(X;G)$ and $a \in C_n(A;G)$ also have that property, i.e. that they are made up only of singular simplices that do not intersect B, then we would be done: indeed, we could then interpret b as an $(n+1)$-chain in $X\backslash B$ and a as an n-chain in $A\backslash B$, so that the relation $c = \partial b + a$ also implies $[c] = 0 \in H_n(X\backslash B, A\backslash B;G)$. As it stands, each of b and a might very well intersect B, but we can now use subdivision to replace them with chains that do not. Indeed, the homology class $[c] \in H_n(X\backslash B, A\backslash B;G)$ does not change if we replace c with $S^m c$ for any $m \geqslant 1$, and since S is a chain map, the relation $c = \partial b + a$ then implies $S^m c = S^m(\partial b) + S^m a = \partial(S^m b) + S^m a$. Choosing m sufficiently large and replacing each of a, b, c with their m-fold subdivisions, we can now assume without loss of generality that all three are decomposable; for $c \in C_n(X\backslash B;G)$ and $a \in C_n(A;G)$ this is not new information since we already assumed them to be contained in $X\backslash B$ or A respectively, but for $b \in C_{n+1}(X;G)$ we can now write

$$b = b_A + b_{X\backslash B} \qquad \text{for some} \qquad b_A \in C_{n+1}(A;G),\; b_{X\backslash B} \in C_{n+1}(X\backslash B;G).$$

The relation $c = \partial b + a$ thus becomes

$$c = \partial b_{X\backslash B} + (\partial b_A + a)\,,$$

and we observe that since c and $\partial b_{X\backslash B}$ are both n-chains in $X\backslash B$, the same must therefore be true for $\partial b_A + a$, meaning it is actually contained in $A\backslash B$. This proves $[c] = 0 \in H_n(X\backslash B, A\backslash B; G)$. $\square$

The remainder of this lecture should be considered optional for now, as it is not needed for the purposes of this semester's course. However, when we study cohomology next semester, we will need a slightly better version of the excision result than Theorem 24.9. One thing you've probably gathered by now is that a chain homotopy is always a useful thing to have, so when one exists, we should take note of it. Theorem 24.9 can be seen as a consequence of the stronger result that the inclusion $i : (X\backslash B, A\backslash B) \hookrightarrow (X, A)$ induces a **chain homotopy equivalence** (*Kettenhomotopieäquivalenz*)

$$i_* : C_*(X\backslash B, A\backslash B; G) \to C_*(X, A; G).$$

In case the meaning of this terminology is not obvious, this means there exists a chain map $\psi : C_*(X, A; G) \to C_*(X\backslash B, A\backslash B; G)$ such that $\psi \circ i_*$ and $i_* \circ \psi$ are each chain homotopic to the identity; we call ψ a **chain homotopy inverse** of i_*.

The following statement turns our previous discussion of subdivision into an actual chain homotopy equivalence that has several applications in the further development of the theory, e.g. we will use it again next semester when we discuss the homology analogue of the Seifert-van Kampen theorem, known as the *Mayer-Vietoris* exact sequence. To understand the statement, it is important to be aware that for any subsets $\mathcal{U}, \mathcal{V} \subset X$, the subgroup $C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G) \subset C_*(X; G)$ is also a chain complex in a natural way. Indeed, the boundary operator on $C_*(X; G)$ maps each of $C_*(\mathcal{U}; G)$ and $C_*(\mathcal{V}; G)$ to themselves, thus it also preserves their sum.

Lemma 24.12. *For any subsets $\mathcal{U}, \mathcal{V} \subset X$ with $X = \mathring{\mathcal{U}} \cup \mathring{\mathcal{V}}$, the inclusion map*

$$j : C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G) \hookrightarrow C_*(X; G)$$

admits a chain homotopy inverse

$$\rho : C_*(X; G) \to C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G)$$

such that $\rho \circ j = \mathbb{1}$, and moreover, there is a chain homotopy $h : C_*(X; G) \to C_{*+1}(X; G)$ of $j \circ \rho$ to the identity such that h vanishes on $C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G)$.

Proof. Let me first point out how one would intuitively wish to prove this, and why it will not work. As observed above, any chain $c \in C_*(X; G)$ can be mapped into $C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G)$ via S^m if the integer m is sufficiently large, so S^m seems like a good candidate for the chain homotopy inverse ρ. The problem however is that we don't know in general how large m needs to be, and in fact the answer depends on the chain c: for any fixed integer m, one can always find a singular n-simplex $\sigma : \Delta^n \to X$ whose boundary is close enough to the boundary of $\mathcal{U}$ or $\mathcal{V}$ so that the m-fold subdivision $S^m(\sigma)$ includes some simplex that is not fully contained in either one. This means that regardless of how large we make m, S^m can never map *all* of $C_*(X; G)$ into $C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G)$, and it will require a bit more cleverness to come up with a candidate for a map ρ that does this. Our approach will be somewhat indirect: instead of writing down ρ, we will first write down a (somewhat naive) candidate for the chain homotopy h in terms of the chain homotopies between S^m and $\mathbb{1}$ for varying values of m. We will then be able to verify that h really is a chain homotopy between $\mathbb{1}$ and something; that so-called "something" will be defined to be ρ, whose further properties we can then verify.

Let $h_1 : C_*(X; G) \to C_{*+1}(X; G)$ denote the chain homotopy provided by Lemma 24.11 for the barycentric subdivision chain map $S : C_*(X; G) \to C_*(X; G)$, i.e. it satisfies $S - \mathbb{1} = \partial h_1 + h_1 \partial$.

We claim that for all integers $m \geqslant 0$, the map

$$h_m := h_1 \sum_{k=0}^{m-1} S^k : C_*(X;G) \to C_{*+1}(X;G)$$

then satisfies

$$S^m - \mathbb{1} = \partial h_m + h_m \partial, \tag{24.2}$$

so h_m is a chain homotopy between S^m and the identity. Note that the case $m = 0$ is included here, with $S^0 = \mathbb{1}$ and $h_0 = 0$, so the claim is trivial in that case, and the definition of h_1 establishes it for $m = 1$. If we now use induction and assume that the claim holds for powers of S up to $m - 1 \geqslant 1$, then since S commutes with ∂,

$$\begin{aligned} S^m - \mathbb{1} &= (S^{m-1} - \mathbb{1})S + (S - \mathbb{1}) = (\partial h_{m-1} + h_{m-1}\partial)S + \partial h_1 + h_1 \partial \\ &= \left(\partial h_1 \sum_{k=0}^{m-2} S^k + h_1 \sum_{k=0}^{m-2} S^k \partial \right) S + \partial h_1 + h_1 \partial = \partial h_1 \sum_{k=1}^{m-1} S^k + h_1 \sum_{k=1}^{m-1} S^k \partial + \partial h_1 + h_1 \partial \\ &= \partial h_1 \sum_{k=0}^{m-1} S^k + h_1 \sum_{k=0}^{m-1} S^k \partial = \partial h_m + h_m \partial. \end{aligned}$$

For any given $\sigma : \Delta^n \to X$, the iterated subdivision maps S^m can be assumed to satisfy

$$S^m(\sigma) \in C_*(\mathcal{U};G) + C_*(\mathcal{V};G) \tag{24.3}$$

if m is large enough, so for each each $n \geqslant 0$ and $\sigma : \Delta^n \to X$, let $m_\sigma \geqslant 0$ denote the smallest integer for which (24.3) holds with $m = m_\sigma$. We can then define a homomorphism $h : C_n(X;G) \to C_{n+1}(X;G)$ for each $n \geqslant 0$ via

$$h(\sigma) := h_{m_\sigma}(\sigma).$$

Let us see whether this is a chain homotopy. We have

$$\begin{aligned} (\partial h + h\partial)(\sigma) &= \partial h_{m_\sigma}(\sigma) + h_{m_\sigma}(\partial\sigma) + (h - h_{m_\sigma})(\partial\sigma) \\ &= (S^{m_\sigma} - \mathbb{1})(\sigma) + (h - h_{m_\sigma})(\partial\sigma) = \left([S^{m_\sigma} + (h - h_{m_\sigma})\partial] - \mathbb{1} \right)(\sigma). \end{aligned}$$

Use this to define $\rho : C_*(X;G) \to C_*(X;G)$ by

$$\rho(\sigma) := S^{m_\sigma}(\sigma) + (h - h_{m_\sigma})(\partial\sigma),$$

so the relation

$$\partial h + h\partial = \rho - \mathbb{1} \tag{24.4}$$

is satisfied. The latter implies that ρ is a chain map since applying ∂ from either the left or right on the left hand side of (24.4) gives $\partial h \partial$, thus on the right hand side we obtain $(\rho - \mathbb{1})\partial = \partial(\rho - \mathbb{1})$. To understand ρ better, we need to observe that each boundary face τ appearing in $\partial\sigma$ satisfies $m_\tau \leqslant m_\sigma$ since m_σ is clearly enough (but need not be the minimal number of) iterations of S to put σ (and therefore also τ) in $C_*(\mathcal{U};G) + C_*(\mathcal{V};G)$. Now if $\sigma \in C_*(\mathcal{U};G) + C_*(\mathcal{V};G)$, then $S^{m_\sigma}(\sigma) = \sigma$ since $m_\sigma = 0$, and the above remarks imply $h(\partial\sigma) = h_0(\partial\sigma) = 0$ as well, thus $\rho(\sigma) = \sigma$ and we conclude

$$\rho \circ j = \mathbb{1}.$$

It remains to show that for all $\sigma : \Delta^n \to X$, $\rho(\sigma)$ is a linear combination of simplices that are each contained in either $\mathcal{U}$ or $\mathcal{V}$. We have $S^{m_\sigma}(\sigma) \in C_*(\mathcal{U};G) + C_*(\mathcal{V};G)$ by the definition of m_σ, so it suffices to inspect the other term $(h - h_{m_\sigma})(\partial\sigma)$. Here again we observe that $\partial\sigma$ is a sum of singular $(n-1)$-simplices τ for which $m_\tau \leqslant m_\sigma$, and

$$(h - h_{m_\sigma})\tau = (h_{m_\tau} - h_{m_\sigma})\tau = -h_1 \sum_{k=m_\tau}^{m_\sigma - 1} S^k(\tau) \in C_n(\mathcal{U}; G) + C_n(\mathcal{V}; G).$$

This last conclusion requires you to recall how h_1 was constructed in the proof of Lemma 24.11: in particular, it maps any simplex that is contained in either $\mathcal{U}$ or $\mathcal{V}$ to a linear combination of simplices that have this same property.

One last detail: the chain homotopy $h : C_*(X; G) \to C_{*+1}(X; G)$ vanishes on $C_*(\mathcal{U}; G) + C_*(\mathcal{V}; G)$ since every singular n-simplex $\sigma : \Delta^n \to X$ with image in either $\mathcal{U}$ or $\mathcal{V}$ satisfies $m_\sigma = 0$, thus $h(\sigma) = h_{m_\sigma}(\sigma) = h_0(\sigma) = 0$. □

Now we can prove the "chain level" result that implies Theorem 24.9.

LEMMA 24.13. *If $A, B \subset X$ are subsets with $\bar{B} \subset \mathring{A}$, then the inclusion $i : (X \backslash B, A \backslash B) \hookrightarrow (X, A)$ induces a chain homotopy equivalence $i_* : C_*(X \backslash B, A \backslash B; G) \to C_*(X, A; G)$.*

PROOF. Consider the quotient chain complex $(C_*(X\backslash B; G) + C_*(A; G))/C_*(A; G)$, which has a natural identification with the group of all finite sums $\sum_i a_i \sigma_i$ with coefficients $a_i \in G$ and singular simplices $\sigma_i : \Delta^n \to X$ that have image in $X\backslash B$ but not contained in A. The point here is that while simplices with $\sigma(\Delta^n) \subset A$ are also generators of $C_*(X\backslash B; G) + C_*(A; G)$, they are all equivalent to zero in the quotient. As it happens, the quotient complex $C_*(X\backslash B, A\backslash B; G) = C_*(X\backslash B; G)/C_*(A\backslash B; G)$ can be described in exactly the same way, with the same set of generators: singular simplices that are contained in $X\backslash B$ but not contained in A. Since the obvious inclusion $C_*(X\backslash B; G) \hookrightarrow C_*(X\backslash B; G) + C_*(A; G)$ sends $C_*(A\backslash B; G)$ into $C_*(A; G)$, it follows that this inclusion descends to a chain map of quotient complexes

$$C_*(X\backslash B, A\backslash B; G) \to (C_*(X\backslash B; G) + C_*(A; G))/C_*(A; G)$$

which is in fact an *isomorphism* of chain complexes, i.e. it has an inverse, which is also a chain map. This is a trivial observation; we have not done anything interesting yet.

But in light of this identification of two quotient chain complexes, it will suffice to prove that the chain map

$$(C_*(X\backslash B; G) + C_*(A; G))/C_*(A; G) \xrightarrow{j} C_*(X; G)/C_*(A; G) = C_*(X, A; G) \tag{24.5}$$

induced on these quotients by the obvious inclusion

$$C_*(X\backslash B; G) + C_*(A; G) \overset{j}{\hookrightarrow} C_*(X; G)$$

is a chain homotopy equivalence. Since $X\backslash\bar{B}$ and $\mathring{A}$ form an open cover of X, Lemma 24.12 provides a chain homotopy inverse for j, namely the map $\rho : C_*(X; G) \to C_*(X\backslash B; G) + C_*(A; G)$, defined in terms of subdivision. That map satisfies $\rho \circ j = \mathbb{1}$, thus ρ restricts to the identity on the subgroup $C_*(A; G) \subset C_*(X; G)$ and therefore descends to a map on quotients going the opposite direction to j in (24.5). It also satisfies $j \circ \rho - \mathbb{1} = \partial h + h\partial$ for a chain homotopy $h : C_*(X; G) \to C_{*+1}(X; G)$ that vanishes on $C_*(A; G)$, thus h also descends to the quotient $C_*(X; G)/C_*(A; G)$ as a chain homotopy $h : C_*(X, A; G) \to C_{*+1}(X, A; G)$ satisfying $j \circ \rho - \mathbb{1} = \partial h + h\partial$ on the quotient complexes. □

REMARK 24.14. We will not need it this semester, but since the notions of chain maps and chain homotopies did not appear in our discussion of simplicial homology, you might wonder if they nonetheless have some role to play in that context. Chain maps arise for instance from *simplicial maps*: given two simplicial complexes $K = (V, S)$ and $K' = (V', S')$, a map $f : V \to V'$ is called a simplicial map if for every simplex σ of K, the images under f of the vertices of σ form the vertices (possibly with repetition) of a simplex of K'. A simplicial map naturally determines a continuous map of the associated polyhedra $|K| \to |K'|$ which maps each n-simplex in $|K|$ linearly to a k-simplex in $|K'|$ for some $k \leqslant n$. It is not hard to show that f also naturally induces a chain map $f_* : C_*(K; G) \to C_*(K'; G)$, defined by sending each n-simplex σ in K to its image k-simplex in K' if $k = n$ and otherwise sending σ to 0. In light of this, Proposition 22.5 implies (unsurprisingly) that any *bijective* simplicial map from K to K' induces an isomorphism of the simplicial homology groups $H_*^{\Delta}(K; G) \to H_*^{\Delta}(K'; G)$. Chain homotopies play an important role when one considers subdivisions of a simplicial complex, e.g. one can adapt the notion of barycentric subdivision so that it naturally associates to any simplicial complex K a larger complex K' with a homeomorphism of $|K'|$ to $|K|$ such that the simplices in K' triangulate the individual simplices of K into smaller pieces. This defines a chain map $S : C_*(K; G) \to C_*(K'; G)$ sending each simplex of K to the linear combination of simplices of K' that triangulate it, and importantly, S turns out to be a chain homotopy equivalence, so it follows from Proposition 24.4 that the induced homomorphism $S_* : H_*^{\Delta}(K; G) \to H_*^{\Delta}(K'; G)$ is an isomorphism. This was historically considered one of the major motivations to believe that simplicial homology depends only on the underlying space $|K|$ and not on the simplicial complex itself (cf. Theorem 21.16). We saw a closely analogous phenomenon in our proof of the excision property above, though in the simplicial context, one usually has to consult some of the older textbooks (e.g. [**Spa95**] is quite nice) to find adequate discussions of such topics.

25. The homology of the spheres, and applications

It is time to put the results of the last few lectures together and compute $H_*(S^n; \mathbb{Z})$. The computation proceeds by induction on the dimension n, making use of the convenient fact that the suspension of S^n is homeomorphic to S^{n+1}. Suspensions, in fact, provide us with our first interesting example of a homotopy equivalence of pairs.

Example 25.1. Recall from Lecture 11 that the **suspension** (*Einhängung*) SX of a space X is defined by gluing together two copies of its cone,

$$SX = C_+X \cup_X C_-X, \tag{25.1}$$

where $C_+X := ([0, 1] \times X)/(\{1\} \times X)$, $C_-X := ([-1, 0] \times X)/(\{-1\} \times X)$, and we identify X with the subset $\{0\} \times X$ in each. Let $p_\pm \in SX$ denote the points at the tips of the two cones, defined by collapsing $\{\pm 1\} \times X$. Then the inclusion

$$(C_+X, X) \hookrightarrow (SX \backslash \{p_-\}, C_-X \backslash \{p_-\})$$

is a homotopy equivalence of pairs. Indeed, one can define a deformation retraction $H : I \times (SX \backslash \{p_-\}) \to SX \backslash \{p_-\}$ by pushing points in $C_-X \backslash \{p_-\}$ continuously upward toward X while leaving C_+X fixed, so that $H(1, \cdot)$ is the identity while $H(0, \cdot)$ retracts $SX \backslash \{p_-\}$ to C_+X and $H(s, \cdot)$ preserves $C_-X \backslash \{p_-\}$ for every $s \in I$. The resulting retraction of pairs $(SX \backslash \{p_-\}, C_-X \backslash \{p_-\}) \to (C_+X, X)$ is a homotopy inverse for the inclusion. Let us spell this out more explicitly in the special case where $X = S^{n-1}$, so SX is then homeomorphic to S^n. The decomposition (25.1) then becomes a splitting of S^n into two hemispheres $\mathbb{D}^n_+ \cong \mathbb{D}^n \cong \mathbb{D}^n_-$ glued along an "equator" homeomorphic to S^{n-1},

$$S^n \cong \mathbb{D}^n_+ \cup_{S^{n-1}} \mathbb{D}^n_-,$$

and our homotopy equivalence of pairs is now the resulting inclusion map

$$(\mathbb{D}^n_+, S^{n-1}) \hookrightarrow (S^n \backslash \{p_-\}, \mathbb{D}^n_- \backslash \{p_-\}),$$

where p_- is now the "south pole," i.e. the center of $\mathbb{D}^n_-$.

The homotopy equivalence in Example 25.1 gives rise to an interesting relationship between $H_*(X;G)$ and $H_*(SX;G)$ for any space X. Ponder the following diagram:

$$\begin{tikzcd}
H_k(X;G) & & H_{k+1}(SX;G) \arrow[d, "\varphi_*"] \\
H_{k+1}(C_+X,X;G) \arrow[u, "\partial_*"] \arrow[r, "i_*"] & H_{k+1}(SX\backslash\{p_-\},C_-X\backslash\{p_-\};G) \arrow[r, "j_*"] & H_{k+1}(SX,C_-X;G)
\end{tikzcd} \tag{25.2}$$

Here ∂_* denotes the connecting homomorphism from the long exact sequence of the pair (C_+X,X), while the maps j_* and φ_* are induced by the obvious inclusions of pairs

$$(SX\backslash\{p_-\},C_-X\backslash\{p_-\}) \xhookrightarrow{j} (SX,C_-X),$$
$$(SX,\varnothing) \xhookrightarrow{\varphi} (SX,C_-X).$$

Since $\{p_-\} \subset C_-X$ is a closed subset in the interior of C_-X, excision (Theorem 24.9) implies that j_* is an isomorphism. We claim that if $k \geqslant 1$, then ∂_* and φ_* are both also isomorphisms. For the first, consider the long exact sequence of the pair (C_+X,X):

$$\ldots \longrightarrow H_{k+1}(C_+X;G) \longrightarrow H_{k+1}(C_+X,X;G) \xrightarrow{\partial_*} H_k(X;G) \longrightarrow H_k(C_+X;G) \longrightarrow \ldots$$

Since C_+X is contractible, homotopy invariance implies that the first and last of these four terms vanish, as $H_n(\{\mathrm{pt}\};G)=0$ for all $n>0$. The sequence thus becomes

$$0 \longrightarrow H_{k+1}(C_+X,X;G) \xrightarrow{\partial_*} H_k(X;G) \longrightarrow 0$$

for each $k \geqslant 1$, so exactness implies that ∂_* is an isomorphism. For φ_*, we instead take an exerpt from the long exact sequence of (SX,C_-X):

$$\ldots \longrightarrow H_{k+1}(C_-X;G) \longrightarrow H_{k+1}(SX;G) \xrightarrow{\varphi_*} H_{k+1}(SX,C_-X;G) \longrightarrow H_k(C_-X;G) \longrightarrow \ldots$$

The contractibility of C_-X again makes the first and last terms vanish if $k \geqslant 1$, leaving

$$0 \longrightarrow H_{k+1}(SX;G) \xrightarrow{\varphi_*} H_{k+1}(SX,C_-X;G) \longrightarrow 0,$$

so that φ_* is also an isomorphism. We have proved:

THEOREM 25.2. *For all spaces X, abelian groups G and integers $k \geqslant 1$, the diagram* (25.2) *defines an isomorphism*

$$S_* = \varphi_*^{-1} \circ j_* \circ i_* \circ \partial_*^{-1} : H_k(X;G) \to H_{k+1}(SX;G).$$

□

EXERCISE 25.3. Show that for any k-cycle $b \in C_k(X;G) \subset C_k(SX;G)$, there exists a pair of $(k+1)$-chains $c_\pm \in C_{k+1}(C_\pm X;G) \subset C_{k+1}(SX;G)$ satisfying

$$\partial c_+ = -\partial c_- = b \tag{25.3}$$

and

$$S_*[b] = [c_+ + c_-]. \tag{25.4}$$

Note that $c_+ + c_- \in C_{n+1}(SX;G)$ is automatically a cycle since $\partial c_+ = -\partial c_-$. Show moreover that (25.4) is satisfied for any pair of chains $c_\pm$ satisfying (25.3).

For the spheres S^n with $n \geqslant 1$, we already know $H_0(S^n;G)$ and $H_1(S^n;\mathbb{Z})$; the former is G because S^n is path-connected (Proposition 22.8), and the latter is the abelianization of $\pi_1(S^n)$ by Theorem 22.10. Since $SS^n \cong S^{n+1}$, we can now compute $H_*(S^n;\mathbb{Z})$ inductively for every $n \geqslant 1$:

THEOREM 25.4. *For every $n \in \mathbb{N}$,*

$$H_k(S^n;\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0, n, \\ 0 & \text{for all other } k. \end{cases}$$

PROOF. Proposition 22.8 gives $H_0(S^n;\mathbb{Z}) \cong \mathbb{Z}$. For $k = n$, $H_n(S^n;\mathbb{Z}) \cong \mathbb{Z}$ follows by an inductive argument starting from $H_1(S^1;\mathbb{Z}) \cong \pi_1(S^1) \cong \mathbb{Z}$ and applying Theorem 25.2. For any $k = 1, \ldots, n-1$, a similar inductive argument starting from $H_1(S^{n-k+1};\mathbb{Z}) = \pi_1(S^{n-k+1}) = 0$ gives $H_k(S^n;\mathbb{Z}) = 0$. For $k > n$, repeatedly applying Theorem 25.2 identifies $H_k(S^n;\mathbb{Z})$ with $H_{k-n}(S^0;\mathbb{Z})$, where $k - n > 0$ and S^0 is a discrete space of two points. But one can easily adapt Exercise 22.9 to prove by direct computation that $H_m(X;G) = 0$ for any $m > 0$ whenever X is a discrete space. □

We can now extend our proof of the Brouwer fixed point theorem to all dimensions. The basic ingredients are the same as before: first, if a map $f : \mathbb{D}^n \to \mathbb{D}^n$ has no fixed point, then we can use it to define a retraction $g : \mathbb{D}^n \to S^{n-1} = \partial\mathbb{D}^n$. In Lecture 10, we used the fundamental group to prove that no such retraction exists when $n = 2$. The argument for this did not require many specific properties of the fundamental group: the key point was just the fact that continuous maps $X \to Y$ induce homomorphisms $\pi_1(X) \to \pi_1(Y)$ in a way that is compatible with composition of maps, and the homology groups have this same property. In particular:

EXERCISE 25.5. Show that if $f : X \to A$ is a retraction to a subset $A \subset X$ with inclusion $i : A \hookrightarrow X$, then for all $n \in \mathbb{Z}$ and abelian groups G, $f_* : H_n(X;G) \to H_n(A;G)$ is surjective, while $i_* : H_n(A;G) \to H_n(X;G)$ is injective.

PROOF OF THE BROUWER FIXED POINT THEOREM. Arguing by contradiction, assume a map $f : \mathbb{D}^n \to \mathbb{D}^n$ without fixed points exists, and therefore also a retraction $g : \mathbb{D}^n \to S^{n-1}$. We may assume $n \geqslant 2$ since the case $n = 1$ follows already from the intermediate value theorem for continuous functions on $[-1, 1]$. By Exercise 25.5, g induces a surjective homomorphism

$$g_* : H_{n-1}(\mathbb{D}^n;\mathbb{Z}) \to H_{n-1}(S^{n-1};\mathbb{Z}).$$

But this is impossible since $H_{n-1}(\mathbb{D}^n;\mathbb{Z}) \cong H_{n-1}(\{\text{pt}\};\mathbb{Z}) = 0$ and $H_{n-1}(S^{n-1};\mathbb{Z}) \cong \mathbb{Z}$. □

Here is another easy application.

THEOREM 25.6. *A topological manifold of dimension n is not also a topological manifold of dimension $m \neq n$.*

PROOF. Let us assume m and n are both at least 2, as the result can otherwise be proved via easier methods. (Hint: removing a point from $\mathbb{R}$ makes it disconnected.) We argue by contradiction and assume M is a manifold with an interior point admitting a neighborhood homeomorphic to $\mathbb{R}^n$ and also a neighborhood homeomorphic to $\mathbb{R}^m$ for $m \neq n$. By choosing a suitable pair of charts and writing down their transition maps, we can produce from this a pair of open neighborhoods of the origin $\Omega_n \subset \mathbb{R}^n$ and $\Omega_m \subset \mathbb{R}^m$ admitting a homeomorphism $f : \Omega_n \to \Omega_m$ with $f(0) = 0$. Choose $\epsilon > 0$ small enough so that f maps the ϵ-ball $B^n_\epsilon(0) \subset \Omega_n$ about the origin into the δ-ball $B^m_\delta(0) \subset \mathbb{R}^m$ for some $\delta > 0$, where the latter is also small enough so that $B^m_\delta(0) \subset \Omega_m$. Now pick a generator

$$A \in H_{n-1}(B^n_\epsilon(0)\backslash\{0\};\mathbb{Z}) \cong H_{n-1}(S^{n-1};\mathbb{Z}) \cong \mathbb{Z}.$$

Since $m \neq n$,

$$H_{n-1}(B^m_\delta(0)\backslash\{0\};\mathbb{Z}) \cong H_{n-1}(S^{m-1};\mathbb{Z}) = 0,$$

so restricting f to a map $B^n_\epsilon(0)\backslash\{0\} \to B^m_\delta(0)\backslash\{0\}$ gives $f_*A = 0 \in H_{n-1}(B^m_\delta(0)\backslash\{0\}; \mathbb{Z})$. But f^{-1} is also defined on $B^m_\delta(0)$, and restricting both f and f^{-1} to maps on punctured neighborhoods with the origin removed, we deduce

$$A = (f^{-1} \circ f)_* A = f^{-1}_* f_* A = 0,$$

which is a contradiction since A was assumed to generate $H_{n-1}(B^n_\epsilon(0)\backslash\{0\}; \mathbb{Z}) \neq 0$. □

26. Axioms, cells, and the Euler characteristic

At this point, I believe I've proved everything that I promised to prove in earlier lectures, so the course *Topologie I* is officially over. Since we nonetheless have a bit of time left, the present lecture is included partly just for fun: none of what it contains should be considered examinable in the current semester, though some of it may provide a useful wider perspective on the material we've previously covered. All of it will also be treated in much more detail in next semester's *Topologie II* course.

The Eilenberg-Steenrod axioms. First a bit of good news: while the proofs of homotopy invariance and excision in Lecture 24 may have seemed somewhat unpleasant, we will hardly ever need to engage in such hands-on constructions via subdivision of simplices in the future. That is because almost everything one actually needs to know in order to use homology in applications follows from a small set of results that we've spent the last few lectures proving. These results form an axiomatic description of general "homology theories," which was first codified by Eilenberg-Steenrod [ES52] and Milnor [Mil62] around the middle of the 20th century. An **axiomatic homology theory** can be thought of as a function

$$(X, A) \mapsto h_*(X, A)$$

that associates to each pair of spaces a sequence of abelian groups $\{h_n(X, A)\}_{n\in\mathbb{Z}}$, and has some additional properties that make it computable for nice spaces and useful for applications in the same way that singular homology is. Identifying each single space X with the pair $(X, \varnothing)$ as usual, one abbreviates

$$h_n(X) := h_n(X, \varnothing).$$

Besides the actual groups $h_n(X, A)$, the theory h_* comes with some additional data: first, it should also associate to each map of pairs $f : (X, A) \to (Y, B)$ a sequence of homomorphisms

$$f_* : h_n(X, A) \to h_n(Y, B), \qquad n \in \mathbb{Z}$$

with the properties that $(f \circ g)_* = f_* \circ g_*$ whenever the composition of f and g makes sense, and the identity map $\mathrm{Id} : (X, A) \to (X, A)$ gives rise to the identity homomorphism $\mathrm{Id}_* = \mathbb{1} : h_n(X, A) \to h_n(X, A)$. Category theory has a technical term for things like this: we call h_* a **functor** from the category of pairs of topological spaces to the category of $\mathbb{Z}$-graded abelian groups. There is one additional piece of data: since the long exact sequences of pairs in singular homology were very useful in the computation of $H_*(S^n)$, we would like to have similar exact sequences for h_*, and one of the ingredients required for this is a sequence of *connecting homomorphisms*

$$\partial_* : h_n(X, A) \to h_{n-1}(A), \qquad n \in \mathbb{Z}.$$

Aside from fitting into an exact sequence as described below, we want these maps to be compatible with the homomorphisms induced on h_* by maps of pairs, in the following sense: any map of pairs $f : (X, A) \to (Y, B)$ restricts to a continuous map $A \to B$, so it induces homomorphisms

$f_* : h_n(X,A) \to h_n(Y,B)$ and $f_* : h_n(A) \to h_n(B)$, which we would like to fit together with ∂_* into the following commutative diagram for each n:

$$\begin{tikzcd}
h_n(X,A) \arrow[r, "\partial_*"] \arrow[d, "f_*"] & h_{n-1}(A) \arrow[d, "f_*"] \\
h_n(Y,B) \arrow[r, "\partial_*"] & h_{n-1}(B)
\end{tikzcd}$$

The fancy category-theoretic term for this condition is "naturality": more specifically, ∂_* defines for each $n \in \mathbb{Z}$ a so-called **natural transformation** from the functor $(X,A) \mapsto h_n(X,A)$ to the functor $(X,A) \mapsto h_n(A) := h_n(A, \varnothing)$. The precise meanings of these terms from category theory will be discussed in the first lecture of next semester's course.

The original list of axioms stated in [ES52] included the properties described above, but they are usually not regarded as actual axioms in modern treatments, since they can instead be summarized with category-theoretic terminology such as "h_* is a functor and ∂_* is a natural transformation". The further conditions we want these things to satisfy are then the following:

- (HOMOTOPY) $f_* : h_*(X,A) \to h_*(Y,B)$ depends only on the homotopy class of $f : (X,A) \to (Y,B)$.
- (EXACTNESS) For the inclusions $i : A \hookrightarrow X$ and $j : (X,\varnothing) \hookrightarrow (X,A)$, the sequence
$$\ldots \longrightarrow h_{n+1}(X,A) \xrightarrow{\partial_*} h_n(A) \xrightarrow{i_*} h_n(X) \xrightarrow{j_*} h_n(X,A) \xrightarrow{\partial_*} h_{n-1}(A) \longrightarrow \ldots$$
is exact.
- (EXCISION) If $B \subset \bar{B} \subset \mathring{A} \subset A \subset X$, then the inclusion $(X \backslash B, A \backslash B) \hookrightarrow (X,A)$ induces an isomorphism $h_*(X \backslash B, A \backslash B) \to h_*(X,A)$.
- (DIMENSION) $h_n(\{\text{pt}\}) = 0$ for all $n \neq 0$. The potentially nontrivial abelian group
$$G := h_0(\{\text{pt}\})$$
is then called the **coefficient group** of h_*.
- (ADDITIVITY) For any collection of spaces $\{X_\alpha\}_{\alpha \in J}$ with inclusion maps $i^\alpha : X_\alpha \hookrightarrow \coprod_{\beta \in J} X_\beta$, the homomorphisms $i^\alpha_* : h_*(X_\alpha) \to h_*(\coprod_\beta X_\beta)$ determine an isomorphism
$$\bigoplus_{\alpha \in J} h_*(X_\alpha) \to h_*\Big(\coprod_{\alpha \in J} X_\alpha\Big).$$

Put together, these properties of an axiomatic homology theory h_* are known as the **Eilenberg-Steenrod axioms**, and they were first written down in [ES52] with the exception of the additivity axiom, which was added later by Milnor [Mil62].[41] We have already done most of the work of proving that for any given abelian group G, the singular homology $H_*(\cdot; G)$ defines an axiomatic homology theory with coefficient group G. The next two exercises fill the remaining gaps in proving this.

EXERCISE 26.1. Assume G is any abelian group and abbreviate the singular homology of a pair (X,A) with coefficients in G by $H_*(X,A) := H_*(X,A;G)$.

(a) Show that the connecting homomorphisms $\partial_* : H_n(X,A) \to H_{n-1}(A)$ in singular homology satisfy naturality, i.e. for any map $f : (X,A) \to (Y,B)$ and every $n \in \mathbb{Z}$, the

---

[41] One can show that for *finite* disjoint unions, the additivity axiom follows from the others—it was thus unnecessary from the perspective of Eilenberg and Steenrod because they were mainly interested in compact spaces, in particular the polyhedra of finite simplicial complexes. The extra axiom becomes important however as soon as the discussion is extended to include noncompact spaces with infinitely many connected components.

diagram

$$\begin{tikzcd} H_n(X,A) \arrow[r, "\partial_*"] \arrow[d, "f_*"] & H_{n-1}(A) \arrow[d, "f_*"] \\ H_n(Y,B) \arrow[r, "\partial_*"] & H_{n-1}(B) \end{tikzcd}$$

commutes.

(b) Deduce that for any map $f : (X,A) \to (Y,B)$, the long exact sequences of (X,A) and (Y,B) in singular homology form the rows of a commutative diagram

$$\begin{tikzcd} \ldots \arrow[r] & H_n(A) \arrow[r] \arrow[d, "f_*"] & H_n(X) \arrow[r] \arrow[d, "f_*"] & H_n(X,A) \arrow[r] \arrow[d, "f_*"] & H_{n-1}(A) \arrow[r] \arrow[d, "f_*"] & \ldots \\ \ldots \arrow[r] & H_n(B) \arrow[r] & H_n(Y) \arrow[r] & H_n(Y,B) \arrow[r] & H_{n-1}(B) \arrow[r] & \ldots \end{tikzcd}$$

Exercise 26.2. Prove directly from the definition of singular homology $H_*(\cdot;G)$ with any coefficient group G that it satisfies the additivity axiom.

If you look again at our computation of $H_*(S^n;\mathbb{Z})$, you'll see that it mostly only used the axioms listed above—I say "mostly" because we did cheat slightly in using the isomorphism $H_1(S^n;\mathbb{Z}) \cong \pi_1(S^n)$, the proof of which is a fairly hands-on argument with singular simplices and does not follow from the axioms. But actually, we could have gotten around this with a little more effort, and it is even possible to compute $H_1(S^n;G)$ for arbitrary coefficient groups G without knowing anything about the fundamental group. The reason we had to appeal to the fundamental group was that Theorem 25.2 is not true for $k = 0$, and it fails for a very specific reason: since H_0 of a contractible space does not vanish, the exact sequences do not always give isomorphisms when this term appears. But there is a formal trick to avoid this problem, called **reduced homology**: it is a variant $\widetilde{H}_*$ of the usual singular homology H_* that fits into all the same exact sequences, but is defined in a slightly more elaborate way so that $\widetilde{H}_n(\{\text{pt}\}) = 0$ for all n, not just for $n \neq 0$. If we had used this, we could have done an inductive argument reducing the homology of every sphere S^n to the homology of S^0, which is the disjoint union of two one-point spaces, so the dimension and additivity axioms then provide the answer. This version of the argument eliminates any need for specifying the coefficients $G = \mathbb{Z}$, and it also works for any axiomatic homology theory, thus giving:

Theorem. *For every $n \in \mathbb{N}$ and any theory h_* satisfying the Eilenberg-Steenrod axioms with coefficient group G,*

$$h_k(S^n) \cong \begin{cases} G & \text{for } k = 0, n, \\ 0 & \text{for all other } k. \end{cases}$$

Now a word of caution: in the last few lectures, we proved two things about singular homology that cannot be deduced merely from the formal properties codified in the Eilenberg-Steenrod axioms, and they are in fact *not true* for arbitrary axiomatic homology theories. One of these was Proposition 22.8, which related H_0 of an arbitrary space X to the set $\pi_0(X)$ of path-components of X via the formula

$$H_0(X;G) \cong \bigoplus_{\pi_0(X)} G. \tag{26.1}$$

This looks at first like it should be related to the additivity axiom: if X is homeomorphic to the disjoint union of its path-components $X_\alpha \subset X$, then additivity gives $H_0(X;G) \cong \bigoplus_\alpha H_0(X_\alpha;G)$, but there is unfortunately nothing in the axioms to imply $H_0(X_\alpha; G) \cong G$ for an arbitrary path-connected space X_α, unless X_α happens to be contractible. There is also a more serious problem, though you may have forgotten about it since we started focusing only on "nice" spaces after Lecture 7: not every space is homeomorphic to the disjoint union of its path-components. Manifolds have this property, and so do locally path-connected spaces in general—the latter follows from a combination of Exercise 7.12, Proposition 7.18 and Theorem 7.19. But not every space is locally path-connected, and no such assumption was imposed on X when we computed $H_0(X; G)$.

Another important result that does not follow from the axioms is Theorem 22.10, on the natural homomorphism

$$\pi_1(X) \to H_1(X; \mathbb{Z}) \tag{26.2}$$

for any path-connected space X, and the isomorphism it induces between $H_1(X; \mathbb{Z})$ and the abelianization of $\pi_1(X)$. Its proof (carried out in Exercise 22.12) similarly required a hands-on examination of the chain complex $C_*(X; \mathbb{Z})$ that underlies the definition of $H_*(X; \mathbb{Z})$. In this context, allow me to point out an odd detail that you may or may not have noticed about the Eilenberg-Steenrod axioms: they never mention any chain complex at all. Homology theories in the sense of Eilenberg-Steenrod need not generally come from chain complexes—in practice, most of them do, though often in less direct ways than singular homology, and one cannot derive from the axioms any direct intuition about the geometric meaning of elements in the groups $h_0(X)$ and $h_1(X)$. Part of the point of the axioms is that for most of the interesting applications of homology, it should suffice to know that a homology theory *exists* and satisfies the right formal properties, because if those properties hold, then one can typically carry out the applications one wants without even knowing how the theory itself is defined. This "highbrow" perspective does not suffice however for computations like (26.1) and (26.2), which are unique to singular homology and its underlying chain complex.

A sketch of Čech homology. Singular homology is not the only theory that satisfies the Eilenberg-Steenrod axioms, though it has been the standard one that people use for over half a century. While the alternatives have gone out of fashion, a few of them do still occasionally resurface in research articles. I would like to give a quick sketch of one of them, if only to demonstrate how two completely different ideas can sometimes lead to invariants that detect more-or-less the same information.

While singular homology tries to understand spaces by viewing singular n-simplices as basic building blocks of n-dimensional objects, the Čech homology theory studies them instead via the combinatorial properties of their open coverings. Suppose in particular that $\mathcal{O} := \{\mathcal{U}_\alpha \subset X\}_{\alpha \in J}$ is an open covering of a space X. One can associate to any such covering an abstract simplicial complex $K_{\mathcal{O}} = (V, S)$, called the **nerve** of the covering: its set of vertices V is the index set J, or equivalently the set of open sets that belong to the covering, and a subset $\sigma := \{\alpha_0, \ldots, \alpha_n\} \subset V$ is defined to be an n-simplex $\sigma \in S$ of the complex $K_{\mathcal{O}}$ if and only if

$$\mathcal{U}_{\alpha_0} \cap \ldots \cap \mathcal{U}_{\alpha_n} \neq \varnothing.$$

This easily satisfies the required conditions for a simplicial complex: each vertex $\alpha \in V$ defines a 0-simplex $\{\alpha\} \in S$ since $\mathcal{U}_\alpha \neq \varnothing$, and each face of $\sigma = \{\alpha_0, \ldots, \alpha_n\} \in S$ is also a simplex in the complex since every nontrivial subcollection of the sets $\mathcal{U}_{\alpha_0}, \ldots, \mathcal{U}_{\alpha_n}$ must still have nonempty intersection. As with all simplicial complexes, $K_{\mathcal{O}}$ gives rise to a topological space, its polyhedron $|K_{\mathcal{O}}|$, but that space need not look at all similar to X: for example, if X is something as simple as S^1, then even if the open covering $\{\mathcal{U}_\alpha\}_{\alpha \in J}$ is finite, the simplicial complex $K_{\mathcal{O}}$ may have arbitrarily large dimension, namely the largest number $n \geqslant 0$ such that $n + 1$ of the sets in the covering have a nonempty intersection.

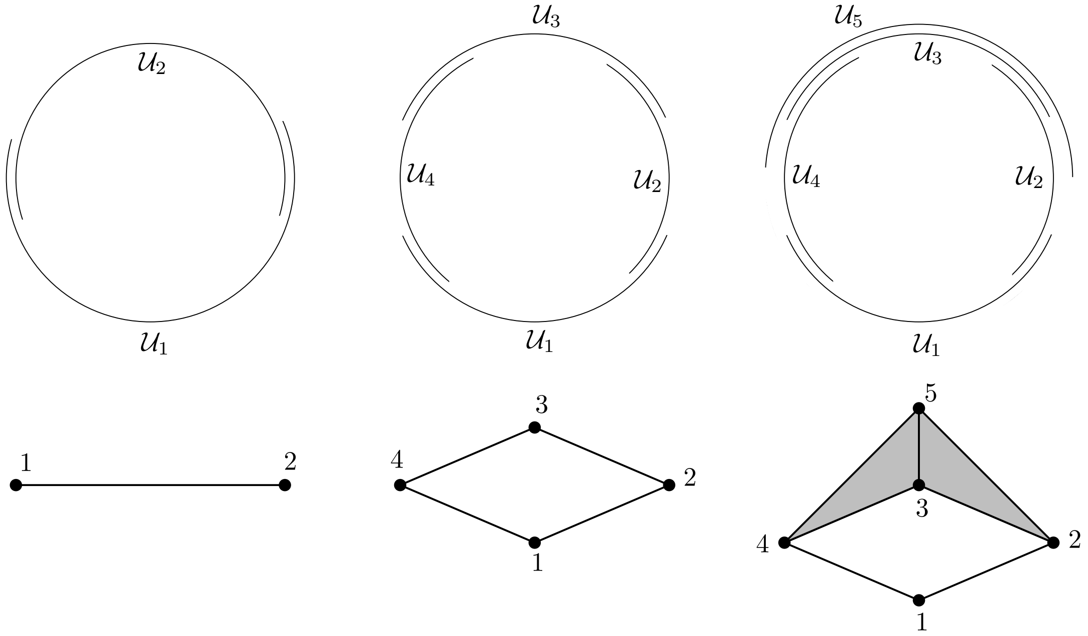


Figure 14. Three examples of open coverings of S^1 and their nerves, with vertices labeled $k \in \{1, 2, 3, 4, 5\}$ in correspondence with the open sets $\mathcal{U}_k \subset S^1$. The rightmost example includes two 2-simplices in addition to vertices and 1-simplices.

The example $X = S^1$ is quite instructive, however, if one compares what $K_{\mathcal{O}}$ looks like for a few simple choices of open coverings. Figure 14 shows three such choices, two of which give rise to 1-dimensional simplicial complexes, and in the third case, the simplicial complex is 2-dimensional. The polyhedra of these three simplicial complexes are all different spaces, none homeomorphic to any of the others, but you may notice that the last two have something in common: they are homotopy equivalent, and not just to each other, but also to the original space, $X = S^1$. The polyhedron in the first example is not homotopy equivalent to S^1, but the other two open coverings also happen to have a nice property that this one does not: in the other two, the intersection sets $\mathcal{U}_{\alpha_0} \cap \ldots \cap \mathcal{U}_{\alpha_n}$ are always contractible, whereas in the first covering, $\mathcal{U}_1 \cap \mathcal{U}_2$ is a disconnected set. Open coverings in which the sets $\mathcal{U}_{\alpha_0} \cap \ldots \cap \mathcal{U}_{\alpha_n}$ are always contractible have a special status: they are called *good* covers, and for sufficiently nice spaces such as smooth manifolds, one can show that every open covering has a refinement that is a good cover. Figure 14 hints at an intriguing general phenomenon: for sufficiently nice open coverings of sufficiently nice spaces X, the nerve of the cover can be viewed as a simplicial model for X itself, up to homotopy type. This suggests that the *simplicial* homology $H_*^{\Delta}(K_{\mathcal{O}}; G)$ of the nerve should encode interesting topological information about X, and that is how Čech homology is defined: for sufficiently nice open coverings $\mathcal{O}$ of X, the **Čech homology** of X with coefficient group G is

$$\check{H}_*(X; G) := H_*^{\Delta}(K_{\mathcal{O}}; G).$$

I am being deliberately vague now, because making this definition more precise would require a discussion of inverse limits and chain homotopy equivalences which we do not have time for right now: in particular, some serious work would be required in order to show that $H_*^{\Delta}(K_{\mathcal{O}}; G)$ up to isomorphism is independent of the choice of (sufficiently nice!) open covering $\mathcal{O}$. The examples on the circle in Figure 14 are intended to convince you that this idea might not be completely outlandish.

Since the definitions of $H_*(X;G)$ and $\check{H}_*(X;G)$ seem very different, it is somewhat remarkable that for a wide class of spaces that includes all compact manifolds, they are isomorphic. One way to explain this is by ignoring the definitions of these two invariants and concentrating instead on their formal properties: after extending Čech homology to an invariant of pairs (X,A) rather than just individual spaces X, one can show (under one or two extra assumptions) that it satisfies the Eilenberg-Steenrod axioms, just like singular homology. As a consequence, any computation that relies only on the formal properties of homology theories—homotopy invariance, excision, long exact sequences and so forth—applies equally well to $H_*(X;G)$ and $\check{H}_*(X;G)$.

It is not true that $H_*(X;G)$ and $\check{H}_*(X;G)$ are always isomorphic, but one has to consider fairly ugly spaces in order to see the difference. A hint of where to look comes from our computation $H_0(X;G) \cong \bigoplus_{\pi_0(X)} G$: as mentioned above, this result does not follow from the axioms. As it turns out, $\check{H}_0(X;G)$ does not care whether the space X is *path*-connected, but cares instead whether it is connected:

Exercise 26.3. Show that if X is a connected space, then for any open cover $\mathcal{O}$ of X, the polyhedron $|K_{\mathcal{O}}|$ of its nerve is path-connected.

Way back in Lecture 7, we saw examples of spaces that are connected but not path-connected. One can deduce from Exercise 26.3 that whenever X is such a space, $\check{H}_0(X;G) \cong G$, but according to (26.1), $H_0(X;G)$ is larger. Using suspensions, one can also derive from this examples of path-connected spaces X for which $\check{H}_1(X;\mathbb{Z})$ is not isomorphic to the abelianization of $\pi_1(X)$. But again: spaces like this are ugly, they are not the kinds of spaces that arise naturally in most applications.

Remark 26.4. In the discussion above, I have swept an uncomfortable fact about $\check{H}_*(X;G)$ under the rug: most versions of Čech homology satisfy *most* of the Eilenberg-Steenrod axioms, but not quite all of them. For technical reasons having to do with the formal properties of inverse limits in homological algebra, $\check{H}_*(X;G)$ does not generally satisfy the exactness axiom unless one restricts to *compact* pairs (X,A) and a restrictive class of coefficient groups G, e.g. any *finite* abelian group or finite-dimensional vector space over a field will do. This shortcoming is one reason why Čech homology has not been used very much in the past half-century. On the other hand, another major topic for next semester's course will be *cohomology*, which is a kind of dualization of homology that has its own closely related set of axioms. The most popular cohomology theory is singular cohomology, but there is also a Čech cohomology theory, which has strictly better formal properties than its undualized counterpart, i.e. it satisfies all of the conditions required for an axiomatic cohomology theory, and even has one or two desirable properties that singular cohomology does not. The ability of Čech cohomology to relate local and global properties of spaces via the combinatorics of their open coverings makes it an essential and frequently used tool in certain branches of mathematics, especially in algebraic geometry.

Cell complexes. We've seen that all axiomatic homology theories are isomorphic on the spaces S^n, though they need not be isomorphic in peculiar examples such as connected spaces that are not path-connected. It is natural to wonder: how large is the class of spaces X for which the Eilenberg-Steenrod axioms completely determine their homologies $h_*(X)$? The spaces with this property happen to be the spaces for which most of the more advanced techniques of algebraic topology have something interesting to say, so they play a starring role in the subject from this point forward.

A plausible first guess for the class of spaces we want to consider would be ***polyhedra***: the topological spaces associated to abstract simplicial complexes. But there is a larger class of spaces called, ***cell complexes*** (or the fancier term "CW-complexes"), which are actually easier to work with and much more general. It is known that all smooth manifolds or simplicial complexes are also cell complexes, and all topological manifolds are at least homotopy equivalent to cell complexes. We saw one concrete example in Lecture 14: when we proved that every finitely presented group occurs as the fundamental group of some compact Hausdorff space (Theorem 14.20), the space we constructed was a wedge of circles with a finite set of disks attached. The general idea of a cell complex is to build up a space inductively as a nested sequence of "skeleta" of various dimensions, where the n-skeleton is always constructed by attaching n-disks to the $(n-1)$-skeleton. In this language, the space constructed in the proof of Theorem 14.20 was a 2-dimensional cell complex, because it had a 1-skeleton (the wedge of circles) and a 2-skeleton (the attached disks). Here is the general definition in the case where there are only finitely many cells.

Definition 26.5. A space X is called a (finite) **cell complex** (*Zellenkomplex*) of dimension n if it contains a nested sequence of subspaces $X^0 \subset X^1 \subset \ldots \subset X^{n-1} \subset X^n = X$ such that:

(1) X^0 is a finite discrete set;
(2) For each $m = 1, \ldots, n$, X^m is homeomorphic to a space constructed by attaching finitely many m-disks $\mathbb{D}^m$ to X^{m-1} along maps $\partial\mathbb{D}^m \to X^{m-1}$.

In general, the collection of m-disks attached to X^{m-1} at each step need not be nonempty; if it is empty, then $X^m = X^{m-1}$, but we implicitly assume $X^n \ne X^{n-1}$ when we call X "n-dimensional".

We call $X^m \subset X$ the m**-skeleton** of X. The definition implies that for each $m = 1, \ldots, n$, there is a finite set $\mathcal{K}_m(X)$ and a so-called **attaching map** $\varphi_\alpha : S^{m-1} \to X^{m-1}$ associated to each $\alpha \in \mathcal{K}_m(X)$ such that

$$X^m \cong \left(\coprod_{\alpha\in\mathcal{K}_m(X)} \mathbb{D}^m \right) \cup_{\varphi_m} X^{m-1},$$

where $\varphi_m : \coprod_{\alpha\in\mathcal{K}_m(X)} \partial\mathbb{D}^m \to X^{m-1}$ denotes the disjoint union of the maps $\varphi_\alpha : S^{m-1} \to X^{m-1}$, each defined on the boundary of the disk indexed by α. As a set, X^m is the union of X^{m-1} with a disjoint union of open disks

$$e^m_\alpha \cong \mathring{\mathbb{D}}^m \quad \text{for each} \quad \alpha \in \mathcal{K}_m(X),$$

called the m**-cells** of the complex. For $m = 0$, we call the discrete points of the 0-skeleton X^0 the **0-cells** and denote this set by $\mathcal{K}_0(X)$.

Since $\Delta^n \cong \mathbb{D}^n$, it is easy to see that polyhedra are also cell complexes: the n-cells are the interiors of the n-simplices, while the n-skeleton is the union of all simplices of dimension at most n and the attaching maps $S^{n-1} \cong \partial\Delta^n \to X^{n-1}$ are each homeomorphisms onto their images. In general, the attaching maps in a cell complex do not need to be injective, they only must be continuous, so while the m-cells e^m_α look like open m-disks, their closures in X might not be homeomorphic to closed disks. For instance, here is an example with an n-cell whose boundary is collapsed to a point, so its closure is not a disk, but a sphere:

Example 26.6. Consider a cell complex that has one 0-cell and no cells of dimensions $1, \ldots, n-1$, so its m-skeleton for every $m < n$ is a one-point space, but there is one n-cell e^n_α attached via the unique map $\varphi_\alpha : S^{n-1} \to \{\text{pt}\}$. The resulting space $X = X^n$ is homeomorphic to S^n.

The **cellular homology** of a cell complex $X = \bigcup_{n \geqslant 0} X^n$ is now defined as follows. Given an abelian coefficient group G, let

$$C_n^{\mathrm{CW}}(X; G) := \bigoplus_{\alpha \in \mathcal{K}_n(X)} G = \left\{ \text{finite sums } \sum_i c_i e^n_{\alpha_i} \;\middle|\; c_i \in G,\ \alpha_i \in \mathcal{K}_n(X) \right\}$$

denote the abelian group of finite linear combinations of generators e^n_α corresponding to the n-cells in the complex, with coefficients in G. A boundary map $\partial : C_n^{\mathrm{CW}}(X; G) \to C_{n-1}^{\mathrm{CW}}(X; G)$ is determined by the formula

$$\partial e^n_\alpha = \sum_{\beta \in \mathcal{K}_{n-1}(X)} [e^{n-1}_\beta : e^n_\alpha] e^{n-1}_\beta,$$

where the **incidence numbers** $[e^{n-1}_\beta : e^n_\alpha] \in \mathbb{Z}$ are determined as follows. For each $\alpha \in \mathcal{K}_n(X)$ and $\beta \in \mathcal{K}_{n-1}(X)$, let

$$X_\beta := X^{n-1} / (X^{n-1} \backslash e^{n-1}_\beta),$$

i.e. it is a space obtained by collapsing everything in the $(n-1)$-skeleton except for the individual cell e^{n-1}_β to a point. Since e^{n-1}_β is an open $(n-1)$-disk with a canonical homeomorphism to $\mathring{\mathbb{D}}^{n-1}$, there is a canonical homeomorphism

$$X_\beta = \mathbb{D}^{n-1} / \partial \mathbb{D}^{n-1} \cong S^{n-1}.$$

There is also a quotient projection $q : X^{n-1} \to X_\beta$, so composing this with the attaching map $\varphi_\alpha : S^{n-1} \to X^{n-1}$ gives a map between two $(n-1)$-dimensional spheres

$$q \circ \varphi_\alpha : S^{n-1} \to X_\beta \cong S^{n-1}.$$

This induces a homomorphism

$$\mathbb{Z} \cong H_{n-1}(S^{n-1}; \mathbb{Z}) \xrightarrow{(q \circ \varphi_\alpha)_*} H_{n-1}(X_\beta; \mathbb{Z}) \cong \mathbb{Z},$$

and all homomorphisms $\mathbb{Z} \to \mathbb{Z}$ are of the form $x \mapsto dx$ for some $d \in \mathbb{Z}$. The integer d appearing here is called the **degree** of $q \circ \varphi_\alpha$, and that is how we define the incidence number:

$$[e^{n-1}_\beta : e^n_\alpha] := \deg(q \circ \varphi_\alpha).$$

Strictly speaking, this definition only makes sense for $n \geqslant 2$ since our computation of the homology of spheres does not apply to S^0, but this is a minor headache that can easily be fixed with an extra definition, as in simplicial homology.

It would take a lot more time than we have right now to explain why this definition of ∂ is the right one, and why it implies $\partial^2 = 0$ in particular. But if you are willing to accept that for now, then we can define the **cellular homology** (*zelluläre Homologie*) groups

$$H_n^{\mathrm{CW}}(X; G) := H_n \left(C_*^{\mathrm{CW}}(X; G), \partial \right),$$

and we can almost immediately carry out a surprisingly easy computation:

EXAMPLE 26.7. The cell decomposition of S^n in Example 26.6 gives

$$H_k^{\mathrm{CW}}(S^n; G) \cong \begin{cases} G & \text{for } k = 0, n, \\ 0 & \text{for all other } k. \end{cases}$$

Indeed, for $n \geqslant 2$ we can see this without doing any work, because $C_0^{\mathrm{CW}}(S^n; G) \cong C_n^{\mathrm{CW}}(S^n; G) \cong G$ are the only nontrivial chain groups, so ∂ simply vanishes and the homology groups are the chain groups. For $n = 1$ you need a little bit more information that I haven't given you, but one can show also in this case that $\partial = 0$, so the result is the same.

In reality, cellular homology is not a *new* homology theory as such, it is just an extremely efficient way of computing any axiomatic homology theory for spaces that are nice enough to have cell decompositions. The following result has been the main tool used for computations of singular homology for most of its history, and it implies in particular the fact that *simplicial* homology is a topological invariant (cf. Theorem 21.16). We will work through a complete proof next semester, and the first step in that proof will be the computation of $h_*(S^n)$.

Theorem. *For any cell complex X and any axiomatic homology theory h_* with coefficient group G, $H_*^{\mathrm{CW}}(X;G) \cong h_*(X)$.*

This theorem is the real reason why homology is considered one of the "easier" invariants to work with in algebraic topology: for most of the spaces that arise in practice, and all compact manifolds in particular, $H_*(X)$ can be computed after replacing the unmageably large singular chain complex with the cellular chain complex, which is *finitely* generated. Having only finitely many generators means that in principle, one can always just feed all the information from the chain complex into a computer program, then press a button and get an answer.

The Euler characteristic. Here is a remarkable application of cellular homology. To make our lives algebraically a bit easier, let's choose the coefficient group G to be a field $\mathbb{K}$, e.g. $\mathbb{Q}$ or $\mathbb{R}$ will do. This has the advantage of making our chain complexes naturally into vector spaces over $\mathbb{K}$, and the boundary maps are $\mathbb{K}$-linear, so the homology groups are also $\mathbb{K}$-vector spaces. Whenever $H_*(X;\mathbb{K})$ is finite dimensional, we then define the **Euler characteristic** of X as the integer

$$\chi(X) := \sum_{n=0}^{\infty} (-1)^n \dim_{\mathbb{K}} H_n(X;\mathbb{K}) \in \mathbb{Z}.$$

Although each individual term $\dim_{\mathbb{K}} H_n(X;\mathbb{K})$ may in general depend on the choice of field $\mathbb{K}$, one can show that their alternating sum does not.[42] This fact admits a purely algebraic proof, but if X is a finite cell complex, then it also follows from the following much more surprising observation. It is not difficult to prove that whenever (C_*, ∂) is a finite-dimensional chain complex of $\mathbb{K}$-vector spaces, the alternating sum of the dimensions of its homology groups can be computed without computing the homology at all: in fact,

$$\sum_{n\in\mathbb{Z}} (-1)^n \dim_{\mathbb{K}} H_n(C_*, \partial) = \sum_{n\in\mathbb{Z}} (-1)^n \dim_{\mathbb{K}} C_n. \tag{26.3}$$

This follows essentially from the fact that for each $n \in \mathbb{Z}$, writing $Z_n := \ker \partial_n \subset C_n$ and $B_n := \operatorname{im} \partial_{n+1} \subset C_n$, the map $\partial_n : C_n \to C_{n-1}$ descends to an isomorphism $C_n/Z_n \to B_{n-1}$, implying

$$\dim_{\mathbb{K}} C_n - \dim_{\mathbb{K}} Z_n = \dim_{\mathbb{K}} B_{n-1}.$$

Since $H_n(C_*, \partial) = Z_n/B_n$, we also have $\dim_{\mathbb{K}} H_n(C_*, \partial) = \dim_{\mathbb{K}} Z_n - \dim_{\mathbb{K}} B_n$, so combining these two relations and adding things up with alternating signs produces lots of cancelations leading to (26.3). Now apply this to the cellular chain complex, in which each $C_n^{\mathrm{CW}}(X;\mathbb{K})$ is a $\mathbb{K}$-vector space whose dimension is the number of n-cells in the complex. What we learn is that we don't need to know anything about homology in order to compute $\chi(X)$—all we have to do is count cells and add up the counts with signs. The isomorphism $H_*(X;\mathbb{K}) \cong H_*^{\mathrm{CW}}(X;\mathbb{K})$ now implies that the result of this counting game only depends on the space, and not on our choice of how to decompose it into cells:

---

[42] One can also define $\chi(X)$ using integer coefficients in terms of the *ranks* of the abelian groups $H_n(X;\mathbb{Z})$. This is one of the algebraic details I wanted to avoid by using field coefficients.

THEOREM. *For any finite cell complex X,*

$$\chi(X) = \sum_{n=0}^{\infty} (-1)^n \, (\textit{the number of } n\textit{-cells})\,.$$

In particular this applies to simplicial complexes, e.g. if you build a 2-sphere by gluing together triangles along common edges, then no matter how you do it or how many triangles are involved, the number of triangles minus the number of glued edges plus the number of glued vertices will always be

$$\chi(S^2) = \dim_{\mathbb{R}} H_0(S^2;\mathbb{R}) - \dim_{\mathbb{R}} H_1(S^2;\mathbb{R}) + \dim_{\mathbb{R}} H_2(S^2;\mathbb{R}) = 1 - 0 + 1 = 2.$$

It is not much harder to work out the result for Σ_g with any $g \geqslant 0$: the answer is

$$\chi(\Sigma_g) = 2 - 2g,$$

and off the top of my head, I can think of two completely different ways to prove this by decomposing Σ_g into cells and counting them with signs: regardless of the choices in the decomposition, the answer will always be the same. Go ahead. Try it.

Second semester (Topologie II)

27. Categories and functors

The general approach of algebraic topology is to associate to each topological space some algebraic object that can be used to tell "different" spaces apart. One important example we saw last semester was the fundamental group π_1, which assigns to every pair (X, p) consisting of a topological space X with a choice of base point $p \in X$ a group $\pi_1(X, p)$. Another—which will play a major role in this course from the next lecture onward—is singular homology H_*, which assigns to each space X a whole sequence of abelian groups $H_n(X)$ indexed by the nonnegative integers $n \geqslant 0$. It is reasonable to think of these in some sense as "functions" with domains consisting of the collection of all topological spaces (possibly with extra data such as a base point), and targets consisting of the collection of all groups. The first semester of this course did not yet develop the right language to make this notion of a "function" precise, so it is time to do so now.

27.1. Some remarks on set theory. One reason why π_1 cannot actually be called a "function" is that its domain, strictly speaking, is not a set (*Menge*). I encourage you to skip the rest of this paragraph if you are not interested in the finer points of axiomatic set theory or the classic set-theoretic paradoxes... but for those who are still reading, let us agree that there is no such thing as the "set of all topological spaces". Indeed, every set can be made into a topological space by assigning it the discrete topology, so if one can talk about the set of all topological spaces, then one must also be able to talk about the set of all sets, and it is a short step from there to the "set of all sets that do not contain themselves"—at which point we may find ourselves asking whether that particular set contains itself, and promptly jumping off the nearest bridge. The architects of abstract set theory dealt with this dilemma by coming up with a set of axioms that tell you how to construct new sets from old ones, together with a short list of examples of sets (e.g. the empty set) whose existence clearly needs to be assumed, and insisting that *only* collections of objects that arise by applying the given axioms to the given examples should be called *sets*. Of course, we do sometimes also need to discuss collections of objects that do not arise from the axioms of set theory, and the collection of all topological spaces is an example. Such collections are generally called (proper) **classes** (*Klassen*), but since I do not wish to go any further into the subtleties of set theory in this course, I shall continue to refer to them via the informal word **collections**. You should just keep in mind that while such things can be defined, they are not considered equivalent to sets, and thus cannot be used for all the same purposes that sets can: in particular, an arbitrary "collection" cannot serve as the domain of a function according to the standard definitions. This doesn't make it impossible to define something that intuitively resembles a function on the collection of all topological spaces—it only means that when we define such an object, we are not strictly allowed to call it a "function". This problem is easy to solve: we shall simply call it something else.

27.2. Categories. Leaving set theory aside, we now introduce some basic notions from category theory. As the examples below should make clear, a category can often be thought of as an answer to the question, "which field of mathematics are we working in?"

Definition 27.1. A **category** (*Kategorie*) $\mathscr{C}$ consists of the following data:

- A collection (i.e. class) $\mathrm{Ob}_{\mathscr{C}}$, whose elements are called the **objects** (*Objekte*) of $\mathscr{C}$;
- For each $X, Y \in \mathrm{Ob}_{\mathscr{C}}$ a set $\mathrm{Hom}_{\mathscr{C}}(X, Y)$, which we shall often abbreviate as
$$\mathrm{Hom}(X, Y) := \mathrm{Hom}_{\mathscr{C}}(X, Y)$$
when there is no danger of confusion, whose elements are called the **morphisms from** X **to** Y (*Morphismen von X nach Y*). For each $X \in \mathrm{Ob}_{\mathscr{C}}$, there is a distinguished[43] element $\mathrm{Id}_X \in \mathrm{Hom}(X, X)$ called the **identity morphism** of X;
- For each $X, Y, Z \in \mathrm{Ob}_{\mathscr{C}}$, a function
$$\mathrm{Hom}(X, Y) \times \mathrm{Hom}(Y, Z) \to \mathrm{Hom}(X, Z) : (f, g) \mapsto g \circ f \tag{27.1}$$
such that $(f \circ g) \circ h = f \circ (g \circ h)$, and whenever two of the objects match and Id denotes the corresponding distinguished morphism, $f \circ \mathrm{Id} = f = \mathrm{Id} \circ f$.

Notation. For a category $\mathscr{C}$, we will often abuse notation and use the symbol $\mathscr{C}$ to indicate not only the category itself but also its collection of objects, hence
$$X \in \mathscr{C} \qquad \text{actually means} \qquad X \in \mathrm{Ob}_{\mathscr{C}}\,.$$
A morphism $f \in \mathrm{Hom}(X, Y)$ from X to Y will often be denoted with the same arrow notation that is standard for maps between sets, so
$$f : X \to Y \quad \text{or} \quad X \xrightarrow{f} Y \qquad \text{actually means} \qquad f \in \mathrm{Hom}(X, Y).$$
The notation $\mathrm{Hom}(X, Y)$ for a set of morphisms is inspired by Example 27.5 below and similar algebraic examples, in which morphisms are actually homomorphisms respecting given algebraic structures. One also often sees this set denoted by $\mathrm{Mor}(X, Y)$ or $\mathscr{C}(X, Y)$, though we will not use that notation here.

Example 27.2. The category Top has $\mathrm{Ob}_{\mathsf{Top}} = \{\text{topological spaces}\}$ and $\mathrm{Hom}(X, Y) = \{f : X \to Y \mid f \text{ a continuous map}\}$, with Id_X defined for each space X as the identity map and the function (27.1) defined as the usual composition of maps. This defines a category since the identity map is always continuous and the composition of two continuous maps is also continuous. In accordance with the notation convention described above, the statement
$$X \in \mathsf{Top}$$
thus means that X is a topological space.

Example 27.3. The category Set has $\mathrm{Ob}_{\mathsf{Set}} = \{\text{sets}\}$ and $\mathrm{Hom}(X, Y) = \{f : X \to Y\}$, meaning that morphisms are simply maps between sets, with no continuity requirement since there is no topology.

Example 27.4. The objects of Diff are the smooth finite dimensional manifolds, and its morphisms are smooth maps. (As in Example 27.2, the identity is always smooth and the composition of two smooth maps is smooth.)

Example 27.5. The category Grp has $\mathrm{Ob}_{\mathsf{Grp}} = \{\text{groups}\}$, with $\mathrm{Hom}(G, H)$ defined as the set of all group homorphisms $G \to H$ for each $G, H \in \mathsf{Grp}$.

---

[43] The word "distinguished" appears here because part of the structure of the category $\mathscr{C}$ is the knowledge of which morphism should be called "Id_X" for each object X. If we simply required the existence of a morphism that satisfies the conditions stated in the third bullet point, then there might be more than one such element and we would not know which one to call Id_X. But the structure of $\mathscr{C}$ requires each set $\mathrm{Hom}(X, X)$ to contain a specific element that carries that name; there might in theory exist additional morphisms that have the same properties, but only one is called Id_X.

EXAMPLE 27.6. There is a **subcategory** (*Unterkategorie*) Ab of Grp whose objects consist of all *abelian* groups, with morphisms defined the same way as in Grp.

The examples above might give you the impression that in every category, a morphism is just a map that may be required to satisfy some specific properties. But nothing in Definition 27.1 says either that an object must be a kind of set or that a morphism is a map. Here is an example in which the objects are still sets, but the morphisms are *equivalence classes* of maps.

EXAMPLE 27.7. Let hTop denote the category whose objects are the same as in Top, but with $\operatorname{Hom}(X,Y)$ defined as the set of homotopy classes of continuous maps $X \to Y$ and $\operatorname{Id}_X \in \operatorname{Hom}(X,X)$ as the homotopy class of the identity map. The function (27.1) is defined in terms of the usual composition of continuous maps $f : X \to Y$ and $g : Y \to Z$ by

$$[g] \circ [f] := [g \circ f].$$

(Exercise: check that this is well defined!) We call hTop the **homotopy category** of topological spaces.

For some interesting examples in which objects are not sets and the function (27.1) has nothing to do with composition of maps, see Exercises 27.3 and 27.4.

DEFINITION 27.8. In any category, a morphism $f \in \operatorname{Hom}(X,Y)$ is called an **isomorphism** (*Isomorphismus*) if there exists a morphism $f^{-1} \in \operatorname{Hom}(Y,X)$ such that $f^{-1} \circ f = \operatorname{Id}_X$ and $f \circ f^{-1} = \operatorname{Id}_Y$. If an isomorphism exists in $\operatorname{Hom}(X,Y)$, we say that the objects X and Y are **isomorphic** (*isomorph*).

According to this definition, the word "isomorphism" no longer has a strictly algebraic meaning, but will mean whatever is considered to be the notion of "equivalence" in whichever category we are working with. Let's run through the list: an isomorphism in Top is a homeomorphism, in Set it is simply a bijection, in Diff a diffeomorphism, and in Grp or Ab it is the usual notion of group isomorphism. The most interesting case so far is hTop: two objects in hTop are isomorphic if and only if they are homotopy equivalent!

The proof of the following is an easy exercise in applying the axioms of a category:

PROPOSITION 27.9. *For any isomorphism $f : X \to Y$ between two objects of a category, the inverse morphism $f^{-1} : Y \to X$ is unique.* □

REMARK 27.10. It is possible to relax Definition 27.1 by allowing $\operatorname{Hom}(X,Y)$ for each $X, Y \in \mathscr{C}$ to be an arbitrary class rather than a set, in which case we are not strictly allowed to call the composition map $\operatorname{Hom}(X,Y) \times \operatorname{Hom}(Y,Z) \to \operatorname{Hom}(X,Z)$ a "function," but the definition still makes sense. In this more general framework, the notion described in Definition 27.1 with morphisms forming sets instead of proper classes is called a **locally small** category. All of the categories we deal with in this course will be locally small, and it takes some nontrivial effort to think up an example of one that is not, so we will not worry about this level of generality any further.

27.3. Functors. The next definition gives us a way of relating two categories to each other. As inspiration, you can think of π_1, a "function" that associates groups to pointed topological spaces, and in fact does so in a way that makes the groups into *topological invariants*. This results mainly from the fact that continuous maps of spaces induce homomorphisms between the corresponding fundamental groups, implying in particular that homeomorphisms induce group isomorphisms. The notion of a functor is meant as a form of abstract packaging for this idea.

DEFINITION 27.11. Given two categories $\mathscr{C}$ and $\mathscr{D}$, a **functor** (*Funktor*) $\mathcal{F} : \mathscr{C} \to \mathscr{D}$ from $\mathscr{C}$ to $\mathscr{D}$ assigns to each object $X \in \mathscr{C}$ an object $\mathcal{F}(X) \in \mathscr{D}$ and to each morphism $f \in \operatorname{Hom}(X,Y)$ between any two objects $X, Y \in \mathscr{C}$ a morphism $\mathcal{F}(f) \in \operatorname{Hom}(\mathcal{F}(X), \mathcal{F}(Y))$ such that:

(1) $\mathcal{F}(\mathrm{Id}_X) = \mathrm{Id}_{\mathcal{F}(X)}$ for all $X \in \mathscr{C}$;
(2) $\mathcal{F}(f \circ g) = \mathcal{F}(f) \circ \mathcal{F}(g)$ for all $g \in \mathrm{Hom}(X, Y)$ and $f \in \mathrm{Hom}(Y, Z)$, $X, Y, Z \in \mathscr{C}$.

Example 27.12. Denote by Top_* the category whose objects are the **pointed spaces** (X, p), i.e. a topological space X together with a point $p \in X$, with morphisms defined as continuous **pointed maps**, also known as **base point preserving** maps,

$$\mathrm{Hom}((X, p), (Y, q)) := \left\{ f : X \to Y \;\middle|\; f \text{ continuous and } f(p) = q \right\}.$$

The fundamental group then defines a functor $\pi_1 : \mathsf{Top}_* \to \mathsf{Grp}$; indeed, it associates to each pointed space (X, p) the group $\pi_1(X, p)$ and to each pointed map $f : (X, p) \to (Y, q)$ the group homomorphism

$$\pi_1(f) := f_* : \pi_1(X, p) \to \pi_1(Y, q)$$

such that Id_* is the identity homomorphism and $(f \circ g)_* = f_* \circ g_*$.

Example 27.13. There is an obvious functor $\mathsf{Top} \to \mathsf{hTop}$ that sends each object $X \in \mathsf{Top}$ to itself and sends each continuous map $f : X \to Y$ to its homotopy class. This is sometimes called a **forgetful functor**, since it is defined by forgetting some (but not all) of the information carried by the morphisms in Top, i.e. it forgets the actual maps $X \to Y$, but remembers their homotopy classes.

Example 27.14. The fundamental group also defines a functor $\pi_1 : \mathsf{hTop}_* \to \mathsf{Grp}$ where hTop_* is defined to have the same objects as Top_*, but with $\mathrm{Hom}((X, p), (Y, q))$ defined as the set of pointed homotopy classes of maps $(X, p) \to (Y, q)$. (See Theorem 8.11 in Lecture 8 from last semester.) A slightly fancier way to say this is that the functor $\pi_1 : \mathsf{Top}_* \to \mathsf{Grp}$ in Example 27.12 is the composition of two functors

$$\begin{tikzcd}
\mathsf{Top}_* \arrow[r] \arrow[rr, bend left, "\pi_1"] & \mathsf{hTop}_* \arrow[r, "\pi_1"'] & \mathsf{Grp}
\end{tikzcd},$$

in which the first is the pointed analogue of the forgetful functor described in Example 27.13. We say in this situation that the functor $\pi_1 : \mathsf{Top}_* \to \mathsf{Grp}$ **descends** to the (pointed) homotopy category hTop_*.

We will later encounter several algebraic constructions and related topological invariants that satisfy most of the conditions of a functor, but differ in one crucial respect: the morphisms they induce go *the other way.* In practice, this phenomenon often arises from the algebraic notion of dualization, and we'll give an example of this kind immediately after the definition.

Definition 27.15. Given two categories $\mathscr{C}$ and $\mathscr{D}$, a **contravariant functor** (*kontravarianter Funktor*) $\mathcal{F} : \mathscr{C} \to \mathscr{D}$ from $\mathscr{C}$ to $\mathscr{D}$ assigns to each $X \in \mathscr{C}$ some $\mathcal{F}(X) \in \mathscr{D}$ and to each $f \in \mathrm{Hom}(X, Y)$ for $X, Y \in \mathscr{C}$ a morphism $\mathcal{F}(f) \in \mathrm{Hom}(\mathcal{F}(Y), \mathcal{F}(X))$ such that

(1) $\mathcal{F}(\mathrm{Id}_X) = \mathrm{Id}_{\mathcal{F}(X)}$ for all $X \in \mathscr{C}$;
(2) $\mathcal{F}(f \circ g) = \mathcal{F}(g) \circ \mathcal{F}(f)$ for all $g \in \mathrm{Hom}(X, Y)$ and $f \in \mathrm{Hom}(Y, Z)$, $X, Y, Z \in \mathscr{C}$.

A functor that satisfies the original Definition 27.11 instead of Definition 27.15 can be called **covariant** (*kovariant*) when we want to emphasize that it is not contravariant.

Example 27.16. Let $\mathbb{K}$-Vect denote the category of vector spaces over a fixed field $\mathbb{K}$, so $\mathrm{Hom}(V, W) := \mathrm{Hom}_{\mathbb{K}}(V, W)$ is the space of $\mathbb{K}$-linear maps $V \to W$. There is a contravariant functor $\boldsymbol{\Delta} : \mathbb{K}\text{-}\mathsf{Vect} \to \mathbb{K}\text{-}\mathsf{Vect}$ which sends each vector space V to its dual space $\boldsymbol{\Delta}(V) := V^* := \mathrm{Hom}_{\mathbb{K}}(V, \mathbb{K})$ and sends each morphism $A : V \to W$ to its transpose $\boldsymbol{\Delta}(A) := A^* : W^* \to V^*$, defined by $A^*(\lambda)v = \lambda(Av)$ for $\lambda \in W^*$ and $v \in V$. It satisfies the conditions of a functor since $(AB)^* = B^*A^*$ and the transpose of the identity $V \to V$ is the identity $V^* \to V^*$.

Remark 27.17. It is possible to avoid Definition 27.15 by instead defining for each category $\mathscr{C}$ the **opposite category** $\mathscr{C}^{\mathrm{op}}$, which has the same collection of objects but reverses the arrows for all morphisms, meaning $\mathrm{Hom}_{\mathscr{C}^{\mathrm{op}}}(X, Y) := \mathrm{Hom}_{\mathscr{C}}(Y, X)$. A contravariant functor $\mathscr{C} \to \mathscr{D}$ is then the same thing as a covariant functor $\mathscr{C}^{\mathrm{op}} \to \mathscr{D}$.

Example 27.18. One can speak of "functors of multiple variables" in much the same way as with functions. It is not difficult to show for instance that on the category Ab of abelian groups and homomorphisms,

$$\mathrm{Hom} : \mathsf{Ab} \times \mathsf{Ab} \to \mathsf{Ab}$$

defines a functor that is contravariant in the first variable and covariant in the second, assigning to each pair of abelian groups (G, H) the group $\mathrm{Hom}(G, H)$ of homomorphisms $G \to H$.

27.4. Natural transformations. We have one more piece of abstract language to add to this story before we can get back to studying topology. You've often seen the words "natural" or "naturally" appearing in statements of theorems, in order to emphasize that something does not depend on any arbitrary choices. In category theory, these words can be given a precise definition.

Definition 27.19. Given two covariant functors $\mathcal{F}, \mathcal{G} : \mathscr{C} \to \mathscr{D}$, a **natural transformation** (*natürliche Transformation*) T from $\mathcal{F}$ to $\mathcal{G}$ associates to each $X \in \mathscr{C}$ a morphism $T_X \in \mathrm{Hom}(\mathcal{F}(X), \mathcal{G}(X))$ such that for all $X, Y \in \mathscr{C}$ and $f \in \mathrm{Hom}(X, Y)$, the following diagram commutes:

$$\begin{tikzcd}
\mathcal{F}(X) \arrow[r, "T_X"] \arrow[d, "\mathcal{F}(f)"] & \mathcal{G}(X) \arrow[d, "\mathcal{G}(f)"] \\
\mathcal{F}(Y) \arrow[r, "T_Y"] & \mathcal{G}(Y)
\end{tikzcd} \tag{27.2}$$

The statement that T is a natural transformation from $\mathcal{F}$ to $\mathcal{G}$ for two functors $\mathcal{F}, \mathcal{G} : \mathscr{C} \to \mathscr{D}$ is sometimes written with the notation

$$T : \mathcal{F} \Rightarrow \mathcal{G}, \qquad \text{or} \qquad \mathcal{F} \overset{T}{\Longrightarrow} \mathcal{G}, \qquad \text{or} \qquad \begin{tikzcd}
\mathscr{C} \arrow[r, bend left=60, "\mathcal{F}", ""{name=U, below}] \arrow[r, bend right=60, "\mathcal{G}"', ""{name=D}] & \mathscr{D} \arrow[Rightarrow, from=U, to=D, "T"]
\end{tikzcd} .$$

A natural transformation between two contravariant functors can be defined analogously.

Remark 27.20. The meaning of commutative diagrams such as (27.2) in an abstract category-theoretical framework should hopefully be obvious: in the case at hand, the diagram means the relation

$$\mathcal{G}(f) \circ T_X = T_Y \circ \mathcal{F}(f),$$

i.e. it specifies that two specific compositions of morphisms give rise to the same morphism $\mathcal{F}(X) \to \mathcal{G}(Y)$. A very large portion of the important definitions and results in category theory can be expressed in terms of commutative diagrams, which make sense due to the axioms of a category, without needing to assume that objects are sets or that morphisms are maps between them.

We will see some nice topological examples of natural transformations in the context of bordism theory in §27.7 below. Here is an algebraic example that you may have heard of before:

Example 27.21. Consider again the category $\mathbb{K}\text{-}\mathsf{Vect}$ of vector spaces over a fixed field $\mathbb{K}$ as in Example 27.16. There is a covariant functor

$$\boldsymbol{\Delta}^2 : \mathbb{K}\text{-}\mathsf{Vect} \to \mathbb{K}\text{-}\mathsf{Vect},$$

assigning to each $V \in \mathbb{K}\text{-}\mathsf{Vect}$ the dual of its dual space $(V^*)^*$. Let $\mathsf{Id} : \mathbb{K}\text{-}\mathsf{Vect} \to \mathbb{K}\text{-}\mathsf{Vect}$ denote the identity functor on $\mathbb{K}\text{-}\mathsf{Vect}$, which sends each object and morphism to itself. There is then a natural transformation from Id to $\boldsymbol{\Delta}^2$ that assigns to every $V \in \mathbb{K}\text{-}\mathsf{Vect}$ a vector space isomorphism $V \to (V^*)^*$; see Exercise 27.5.

Remark 27.22. Whenever a vector space V is finite dimensional, the map $V \to (V^*)^*$ given by the natural transformation in Example 27.21 is an isomorphism, and a large part of the reason why it turns out to define a *natural* transformation is that the definition of this map does not depend on any choices. By contrast, every finite-dimensional vector space is isomorphic to its dual space V^*, but there is no *canonical* way to define such isomorphisms for all vector spaces at once. Notice that since $\mathsf{Id} : \mathbb{K}\text{-}\mathsf{Vect} \to \mathbb{K}\text{-}\mathsf{Vect}$ is a covariant functor while the dualization functor $\boldsymbol{\Delta} : \mathbb{K}\text{-}\mathsf{Vect} \to \mathbb{K}\text{-}\mathsf{Vect}$ from Example 27.16 is contravariant, there is no sensible notion of natural transformations from Id to $\boldsymbol{\Delta}$.

27.5. Bordism groups.

It would be too ambitious to attempt a serious discussion of bordism theory in this course, but there are two good reasons to introduce the basic definitions now. First, they give us some elegant new topological examples of functors besides π_1, including some obviously interesting examples of natural transformations. Second, the geometric idea behind bordism groups will give us motivation for the somewhat less straightforward definition of homology groups in the lectures to come.

Notation. This is a convenient moment to mention a notational convention that will be in force throughout the semester: we abbreviate the compact unit interval by

$$I := [0, 1].$$

This will be the meaning of the symbol I in any context that involves homotopies.

For some initial motivation, you can think of π_1 in the following terms: first, elements of $\pi_1(X)$ are represented by base-point preserving maps $\gamma : S^1 \to X$ defined on a specific closed 1-dimensional manifold, namely the circle S^1. Two such maps $\gamma, \gamma' : S^1 \to X$ represent the same element if there exists a pointed homotopy

$$h : S^1 \times I \to X,$$

between them, so in this situation, the disjoint union $\gamma \amalg \gamma' : S^1 \amalg S^1 \to X$ of the two maps admits a continuous extension to a map $S^1 \times I \to X$, whose domain is a specific compact 2-dimensional manifold with boundary naturally homeomorphic to $S^1 \amalg S^1$. This way of describing homotopies ignores base points, but base points are not important for our present purposes: what's important rather is that we are talking about maps into X defined on compact 2-manifolds bounded by closed 1-manifolds. If you take this picture and ask what happens when you allow the domains to be arbitrary compact manifolds of arbitrary dimension, bordism theory is what you get.

For the following definition, recall that an n-dimensional manifold M is called **closed** if it is compact and the $(n-1)$-dimensional manifold ∂M defined as the boundar of M is empty. We will generally use the term "manifold" as a synonym for "manifold with boundary," so all manifolds M are allowed to have a nonempty boundary ∂M, but we shall make no overriding assumptions about whether ∂M is nonempty unless extra words such as "closed" are included. It is useful to note however that for any manifold M, the boundary ∂M is a manifold whose own boundary is always empty:

$$\partial(\partial M) = \varnothing.$$

Definition 27.23. For a space $X \in \mathsf{Top}$ and an integer $n \geqslant 0$, the nth **unoriented bordism group** $\Omega_n^{\mathrm{O}}(X)$ of X consists of equivalence classes $[(M, \varphi)]$ of pairs (M, φ) in which M is a closed smooth n-manifold and $\varphi : M \to X$ is a continuous map. We call two such pairs (M, φ) and (N, ψ)

equivalent (or **bordant**) if there exists a **bordism** between them, meaning a pair (W, Φ) in which W is a compact smooth $(n+1)$-manifold, $\Phi : W \to X$ is a continuous map, and there exists a diffeomorphism

$$\partial W \cong M \amalg N$$

such that

$$\Phi|_{\partial W} = \varphi \amalg \psi.$$

We make $\Omega_n^{\mathrm{O}}(X)$ into an abelian group by using disjoint unions to define addition, thus

$$[(M, \varphi)] + [(N, \psi)] := [(M \amalg N, \varphi \amalg \psi)],$$

with the additive identity element defined by

$$0 := [(\varnothing, \cdot)] \in \Omega_n^{\mathrm{O}}(X),$$

where the empty set $\varnothing$ is understood as a smooth manifold of arbitrary dimension, and $\cdot$ denotes the unique map $\varnothing \to X$.

A few observations are needed before this definition fully makes sense. First, we should check that the bordism relation described above satisfies the conditions of an equivalence relation: for instance, it is reflexive because for any closed n-manifold M and map $\varphi : M \to X$, the compact $(n+1)$-manifold $M \times I$ and map

$$M \times I \to X : (x, t) \mapsto \varphi(x) \tag{27.3}$$

define a bordism between (M, φ) and itself. The symmetry of the relation is obvious; the most interesting detail is transitivity, which requires some rudimentary knowledge of smooth manifolds and collar neighborhoods, so that two $(n+1)$-manifolds with diffeomorphic boundary components can be glued together along those components to form a new $(n+1)$-manifold. Since this discussion is not intended as a comprehensive introduction to bordism theory, I will leave that detail to your imagination for now. Once the bordism relation is understood, it is straightforward to check that the addition operation defined via disjoint unions is well defined on equivalence classes. The remaining question to answer is why $\Omega_n^{\mathrm{O}}(X)$ is a group, i.e. why every element has an additive inverse. This also comes from the map (27.3), because there is another way to interpret it: the boundary of $M \times I$ is naturally diffeomorphic to the disjoint union of $M \amalg M$ with $\varnothing$, which makes $(M \amalg M, \varphi \amalg \varphi)$ bordant to $(\varnothing, \cdot)$ and thus proves

$$[(M, \varphi)] + [(M, \varphi)] = 0 \in \Omega_n^{\mathrm{O}}(X).$$

This not only makes $\Omega_n^{\mathrm{O}}(X)$ a group, but also gives it an especially simple algebraic structure: all of its nontrivial elements have order 2, so the abelian group $\Omega_n^{\mathrm{O}}(X)$ can also be regarded as a vector space over the field $\mathbb{Z}_2$.

REMARK 27.24. The domains in Definition 27.23 were all assumed to be *smooth* manifolds rather than just *topological* manifolds, but there is an equally sensible variation on this definition that requires only topological manifolds and replaces the word "diffeomorphism" (in the definition of the bordism relation) with "homeomorphism". The main reason to include smoothness in the definition is that methods from differential topology make $\Omega_n^{\mathrm{O}}(X)$ easier to compute than its purely topological counterpart. But for our present purposes, this detail will make no difference at all and can safely be ignored.

The following observation makes Ω_n^{O} into a covariant functor

$$\Omega_n^{\mathrm{O}} : \mathsf{Top} \to \mathsf{Ab},$$

or equivalently (in light of the fact that all nontrivial elements have order two), a functor $\Omega_n^{\mathrm{O}} : \mathsf{Top} \to \mathbb{Z}_2\text{-}\mathsf{Vect}$. Each continuous map $f : X \to Y$ induces a map $f_* : \Omega_n^{\mathrm{O}}(X) \to \Omega_n^{\mathrm{O}}(Y)$ defined by

$$f_*[(M, \varphi)] := [(M, f \circ \varphi)].$$

It is straightforward to check that this map is well defined and is a group homomorphism. It clearly also sends the identity map $X \to X$ to the identity homomorphism $\Omega_n^{\mathrm{O}}(X) \to \Omega_n^{\mathrm{O}}(X)$ and satisfies the relation $(f \circ g)_* = f_* g_*$ for any two continuous maps f, g that are composable. In other words: $\Omega_n^{\mathrm{O}} : \mathsf{Top} \to \mathsf{Ab}$ is a functor.

A less obvious but very useful observation is that $\Omega_n^{\mathrm{O}} : \mathsf{Top} \to \mathsf{Ab}$ *descends* (in the sense of Example 27.14) to the corresponding homotopy category, and thus also defines a functor

$$\Omega_n^{\mathrm{O}} : \mathsf{hTop} \to \mathsf{Ab}.$$

This is an immediate consequence of the following result:

PROPOSITION 27.25. *For any two homotopic maps $f, g : X \to Y$, the induced homomorphisms $f_*, g_* : \Omega_n^{\mathrm{O}}(X) \to \Omega_n^{\mathrm{O}}(Y)$ are identical.*

PROOF. Assume $H : X \times I \to Y$ is a homotopy with $H(\cdot, 0) = f$ and $H(\cdot, 1) = g$. Given $[(M, \varphi)] \in \Omega_n^{\mathrm{O}}(X)$, the map

$$M \times I \to Y : (x, t) \mapsto H(\varphi(x), t)$$

then defines a bordism between $(M, f \circ \varphi)$ and $(M, g \circ \varphi)$, proving $f_*[(M, \varphi)] = g_*[(M, \varphi)] \in \Omega_n^{\mathrm{O}}(Y)$. ☐

27.6. Oriented bordism. In case you had hoped for a more interesting group in which not all nontrivial elements have order 2, there is a remedy: one can add a bit of extra data to the domain manifolds that are used to represent bordism classes, namely an orientation. If you know already what it means to equip a smooth manifold with an orientation, then great—if not, then this is not the place to discuss it, though we will give a detailed treatment of orientations for *topological* manifolds later in this semester. For present purposes, it will suffice to take the following facts about orientations on faith:

(1) Many familiar manifolds such as S^1 and the compact surfaces Σ_g of genus g for each $g \geqslant 0$ are orientable, but not all manifolds are, e.g. the projective plane $\mathbb{RP}^2$ and the Klein bottle are not. More generally, no manifold that contains a Möbius band (or equivalently, that is the connected sum of something with $\mathbb{RP}^2$) can admit an orientation, because the Möbius band contains a loop such that any choice of orientation at one point gets reversed by moving it continuously along the loop.

(2) For every orientation of a manifold M, there is another orientation called the **opposite orientation**, and if M is connected, then it admits exactly two orientations, which are opposites of each other. For an oriented manifold M, we sometimes denote by $-M$ the same manifold with the opposite orientation.[44]

(3) For every manifold M with nonempty boundary, an orientation of M naturally determines an orientation of ∂M, called the **boundary orientation**. The opposite orientation of M then determines the opposite boundary orientation, or in symbols,

$$\partial(-M) = -(\partial M).$$

---

[44] Another popular way of denoting the oriented manifold $-M$ is $\overline{M}$, especially in certain situations where M comes with a canonical choice of orientation. This is true for instance if M is a *complex* manifold, e.g. the complex projective space $\mathbb{CP}^n$, which inherits a canonical orientation from its complex structure, and $\overline{\mathbb{CP}}^n$ then denotes the same real manifold with an orientation opposite to the one determined by the complex structure.

(4) For any two oriented manifolds M, N with $\dim M = m$ and $\dim N = n$, the Cartesian product $M \times N$ is an $(m+n)$-manifold that inherits a **product orientation**, which depends in general on the order of the factors, though only if both m and n are odd. In symbols,

$$N \times M \cong (-1)^{mn}(M \times N),$$

meaning that the obvious diffeomorphism $M \times N \xrightarrow{\cong} N \times M$ is orientation reversing if m and n are both odd, and is otherwise orientation preserving.

(5) For M a 0-manifold (also known as a discrete set with at most countably many points), an orientation is simply a function $M \to \{1, -1\}$, and for any choice of orientation on the unit interval $I = [0,1]$, the boundary orientation assigns opposite signs to the two points of $\partial I = \{0, 1\}$.

Definition 27.26. The nth **oriented bordism group**

$$\Omega_n^{\mathrm{SO}}(X)$$

of a space X is defined by modifying Definition 27.23 as follows: the manifold M in each representative (M, φ) is equipped with an orientation, and the manifold W in an **oriented bordism** (W, Φ) between (M, φ) and (N, ψ) is also oriented and equipped with an orientation-preserving diffeomorphism

$$\partial W \cong -M \amalg N,$$

where ∂W is assumed to carry the boundary orientation.

Let us clarify why reversing the orientation of either M or N in the oriented bordism relation is the right thing to do. For any oriented manifold M, assigning the product orientation to $M \times I$ and then the boundary orientation to $\partial(M \times I)$ gives a natural *orientation-preserving* diffeomorphism

$$\partial(M \times I) \cong -M \amalg M.$$

The trivial homotopy (27.3) thus implies again that the bordism relation satisfies $(M, \varphi) \sim (M, \varphi)$, but in the oriented setting, it does *not* imply $[(M, \varphi)] + [(M, \varphi)] = 0$, so that elements of $\Omega_n^{\mathrm{SO}}(X)$ do not need to have order two. Instead, the additive inverse of any given $[(M, \varphi)] \in \Omega_n^{\mathrm{SO}}(X)$ is obtained by reversing the orientation of M,

$$-[(M, \varphi)] = [(-M, \varphi)] \in \Omega_n^{\mathrm{SO}}(X).$$

Remark 27.27. The letters "O" and "SO" appearing in the notation $\Omega_n^{\mathrm{O}}(X)$ and $\Omega_n^{\mathrm{SO}}(X)$ refer to the orthogonal group $\mathrm{O}(n)$ and special orthogonal group $\mathrm{SO}(n)$ respectively, which makes some sense if you recall that $\mathrm{SO}(n)$ is precisely the subgroup of $\mathrm{O}(n)$ consisting of transformations $\mathbb{R}^n \to \mathbb{R}^n$ that preserve orientation. A fuller explanation of this notation would be too much of a digression for now, but suffice it to say there also exist other versions of bordism groups corresponding to other families of Lie groups that act linearly on Euclidean space, in which the manifold M in representatives (M, φ) of bordism classes is equipped with extra structure respected by those group actions.

The following easy computation (see Exercise 27.7) demonstrates that, indeed, elements of $\Omega_n^{\mathrm{SO}}(X)$ need not have order 2 in general.

Proposition 27.28. *For any space X, there are natural isomorphisms*

$$\Omega_0^{\mathrm{O}}(X) \cong \bigoplus_{\pi_0(X)} \mathbb{Z}_2, \qquad \text{and} \qquad \Omega_0^{\mathrm{SO}}(X) \cong \bigoplus_{\pi_0(X)} \mathbb{Z},$$

i.e. $\Omega_0^{\mathrm{O}}(X)$ is a vector space over $\mathbb{Z}_2$ with a canonical basis in bijective correspondence with the set $\pi_0(X)$ of path-components of X, and $\Omega_0^{\mathrm{SO}}(X)$ is a free abelian group with the same basis. ☐

Beyond the case $n = 0$, computations of $\Omega_n^{\mathrm{O}}(X)$ and $\Omega_n^{\mathrm{SO}}(X)$ are generally doable, but too difficult to attempt before learning about homology and cohomology, which will be the main objectives of this semester's course. One sees a revealing symptom of the difficulty when one tries to compute either of these groups for the simplest possible nonempty topological space, namely a one-point space.

Notation. We will frequently denote by

$$\{*\} \in \mathsf{Top}$$

a topological space consisting of only one point, with the point in this space denoted by

$$* \in \{*\}.$$

The symbol $* \in X$ is sometimes also used to denote the base point of a pointed space $X \in \mathsf{Top}_*$, if it has not been given any other name.

Remark 27.29. We sometimes abuse terminology by speaking of "the" one-point space, but of course one-point spaces are not unique, since the one element in the space can literally be anything, e.g. the sets $\{1\}$ and $\{2\}$ are not identical since $1 \neq 2$, and they are also different from the set whose only element is the banana you ate for breakfast this morning. In light of the fact that set theory has no way of defining a "set of all things," the collection of all possible one-point spaces forms a proper class rather than a set. However, one does have a strong form of uniqueness up to isomorphism in the category Top or Top_*: there exists a unique homeomorphism between any two one-point spaces, and this is why referring to them all as "the" one-point space does not do any harm.

For a one-point space $\{*\}$ and any given manifold M, there is only one possible map $M \to \{*\}$, so elements of the groups[45]

$$\Omega_n^{\mathrm{O}} := \Omega_n^{\mathrm{O}}(\{*\}), \qquad \Omega_n^{\mathrm{SO}} := \Omega_n^{\mathrm{SO}}(\{*\})$$

can be regarded simply as equivalence classes $[M]$ of closed n-manifolds, and the information encoded in these groups is therefore a coarse version of the classification of closed n-manifolds, subject to an equivalence relation in which boundaries of compact manifolds are equated with the empty set. The classification problem up to homeomorphism or diffeomorphism is well understood for manifolds of dimension at most two, but already from dimension three upward, complete classifications are not known, and the problem is not generally considered tractable. From this perspective, it seems slightly surprising that Ω_n^{O} and Ω_n^{SO} can in fact be computed, and the answers are often not difficult to write down, but proving them usually takes quite a bit of work. By the end of this semester, we will at least be able to fill in all the gaps in the following special case:

Proposition 27.30. *The group $\Omega_2^{\mathrm{O}} = \Omega_2^{\mathrm{O}}(\{*\})$ is isomorphic to $\mathbb{Z}_2$, and its unique nontrivial element is the bordism class of the projective plane $\mathbb{RP}^2$.*

Proof sketch. The nontriviality of $[\mathbb{RP}^2] \in \Omega_2^{\mathrm{O}}(\{*\})$ means that $\mathbb{RP}^2$ is not diffeomorphic to the boundary of any compact 3-manifold. If you take this on faith for a moment, the rest of the computation follows easily from the classification of surfaces, as described in Lecture 19 from last semester. Indeed, every closed and *orientable* surface can be presented as the smooth boundary of a compact region in $\mathbb{R}^3$, and thus represents the trivial element in Ω_2^{O}. The closed, connected

---

[45] The symbols Ω_n^{SO} and Ω_n^{O} now each have two possible interpretations, either as functors $\mathsf{Top} \to \mathsf{Ab}$ or as the groups obtained by plugging a one-point space into these functors. It depends on the context.

and non-orientable surfaces, in turn, are all homeomorphic (and in fact also diffeomorphic[46]) to connected sums of N copies of $\mathbb{RP}^2$ for some $N \in \mathbb{N}$. A convenient fact to use in this situation is that for any two closed manifolds M, N of the same dimension n, there exists a compact $(n+1)$-manifold whose boundary is diffeomorphic to the disjoint union of M, N and the connected sum $M \# N$. This can be proved with a picture, and I will leave it as an exercise, but if you need a hint, try looking up some information on *handle attachment* in geometric topology—the key trick is to "attach a 1-handle" to $(M \amalg N) \times I$. With this understood, one now sees that every closed and connected surface is bordant to some disjoint union of copies of $\mathbb{RP}^2$, and therefore so is every closed and disconnected surface.

So, why is $\mathbb{RP}^2$ not the boundary of any compact 3-manifold? This is harder to explain, but it will follow easily from some computations of homological invariants carried out later in this course. In particular, the Poincaré duality isomorphism implies that the Euler characteristic (an integer-valued invariant that is defined for a wide class of topological spaces including all compact manifolds) of every closed odd-dimensional manifold is zero. If $\mathbb{RP}^2$ were the boundary of some compact 3-manifold Y, then by gluing Y to a copy of itself along the boundary, one would obtain a closed 3-manifold

$$X := Y \cup_{\mathbb{RP}^2} Y$$

whose Euler characteristic $\chi(X)$ satisfies $\chi(X) = 2\chi(Y) - \chi(\mathbb{RP}^2) = 2\chi(Y) - 1$, and therefore could not be zero. ☐

27.7. More examples of natural transformations. The bordism groups provide us with some examples of natural transformations that are quite easy to write down. Proving the required naturality property, i.e. that the required diagrams commute, is a straightforward exercise in each case.

EXAMPLE 27.31. For every space X and $n \geqslant 0$, there is an obvious *forgetful* homomorphism

$$\Omega_n^{\mathrm{SO}} \to \Omega_n^{\mathrm{O}} : [(M, \varphi)] \mapsto [(M, \varphi)]$$

defined by forgetting the orientation of the manifold M. Regarding both Ω_n^{SO} and Ω_n^{O} as covariant functors $\mathsf{Top} \to \mathsf{Ab}$, this defines a natural transformation from Ω_n^{SO} to Ω_n^{O}.

EXAMPLE 27.32. Since S^1 is a closed orientable 1-manifold, one can associate to any pointed space (X, p) a map

$$\pi_1(X, p) \xrightarrow{h} \Omega_1^{\mathrm{SO}}(X) : [\gamma] \mapsto [(S^1, \gamma)], \tag{27.4}$$

defined by regarding representatives of elements in $\pi_1(X, p)$ as maps $\gamma : S^1 \to X$. This map is well defined because a homotopy between two maps $\gamma, \gamma' : S^1 \to X$ gives rise to a bordism between (S^1, γ) and (S^1, γ'). The lemma below shows that $h : \pi_1(X, p) \to \Omega_1^{\mathrm{SO}}(X)$ is also a group homomorphism; it is a variation on what is known in homology theory as the *Hurewicz homomorphism*, and we will later see another version of it in that context.

LEMMA 27.33. *For any two base-point preserving loops $\alpha, \beta : S^1 \to X$ and their concatenation $\alpha \cdot \beta : S^1 \to X$, $(S^1, \alpha \cdot \beta)$ is bordant to $(S^1 \amalg S^1, \alpha \amalg \beta)$.*

PROOF. See Exercise 27.8. ☐

---

[46] It is a nontrivial fact that for $n \leqslant 3$ (though emphatically not for $n \geqslant 4$), every topological n-manifold admits a smooth structure, and two smooth n-manifolds are homeomorphic if and only if they are diffeomorphic. For closed surfaces, the easiest way to prove this is probably by reproving the standard classification of surfaces in the smooth category. In fact, this is easier than working only with topological surfaces and continuous maps, because Riemannian geometry makes the existence of triangulations on *smooth* manifolds easier to prove.

Continuing with Example 27.32, the following "naturality" property of the map $h : \pi_1(X,p) \to \Omega_1^{\mathrm{SO}}(X)$ is nearly immediate from the definitions: for any pointed map $f : (X,p) \to (Y,q)$, the diagram

$$\begin{tikzcd}
\pi_1(X,p) \arrow[r, "h"] \arrow[d, "f_*"] & \Omega_1^{\mathrm{SO}}(X) \arrow[d, "f_*"] \\
\pi_1(Y,q) \arrow[r, "h"] & \Omega_1^{\mathrm{SO}}(Y)
\end{tikzcd}$$

commutes. This *almost* amounts to the statement that h defines a natural transformation from π_1 to Ω_1^{SO}, though before we can say this in precise terms, we have a minor bookkeeping issue to deal with, as $\pi_1 : \mathsf{Top}_* \to \mathsf{Grp}$ and $\Omega_1^{\mathrm{SO}} : \mathsf{Top} \to \mathsf{Ab}$ are not functors between exactly the same pairs of categories, strictly speaking. The distinction between Grp and Ab is easy to erase since the latter is a subcategory of the former, i.e. we can equally well regard Ω_1^{SO} as a functor $\mathsf{Top} \to \mathsf{Grp}$. For the distinction between Top_* and Top, the obvious thing to do is define Ω_1^{SO} as a functor $\mathsf{Top}_* \to \mathsf{Grp}$ by composing the usual $\Omega_1^{\mathrm{SO}} : \mathsf{Top} \to \mathsf{Grp}$ with the obvious *forgetful* functor $\mathsf{Top}_* \to \mathsf{Top}$, replacing each pointed space (X,p) with the unpointed space X. With this understood, the commuting diagram above shows that h defines a natural transformation from π_1 to Ω_1^{SO} if both are regarded as functors $\mathsf{Top}_* \to \mathsf{Grp}$. For a slightly different variation, we could observe that since $\Omega_1^{\mathrm{SO}}(X)$ is abelian, the map $h : \pi_1(X,p) \to \Omega_1^{\mathrm{SO}}(X)$ always vanishes on the commutator subgroup

$$[\pi_1(X,p), \pi_1(X,p)] \subset \pi_1(X,p),$$

and thus descends to a well-defined homomorphism on the abelianization of the fundamental group,

$$\pi_1(X,p)/[\pi_1(X,p), \pi_1(X,p)] \xrightarrow{h} \Omega_1^{\mathrm{SO}}(X).$$

By now you should be unsurprised to learn that *abelianization* can also be regarded as a functor

$$\mathsf{Grp} \xrightarrow{\mathsf{ab}} \mathsf{Ab} : G \mapsto \mathsf{ab}(G) := G\Big/[G,G],$$

and we can then also regard h as a natural transformation from $\mathsf{ab} \circ \pi_1 : \mathsf{Top}_* \to \mathsf{Ab}$ to $\Omega_{\mathrm{SO}}^1 : \mathsf{Top}_* \to \mathsf{Ab}$.

Example 27.34. Let $\Omega_n^\bullet : \mathsf{Top} \to \mathsf{Ab}$ denote either the unoriented or oriented bordism functor. For any two spaces X, Y and integers $m, n \geqslant 0$, one can define a product operation

$$\begin{aligned} \Omega_m^\bullet(X) \otimes \Omega_n^\bullet(Y) &\xrightarrow{\times} \Omega_{m+n}^\bullet(X \times Y), \\ [(M,\varphi)] \otimes [(N,\psi)] &\longmapsto [(M,\varphi)] \times [(N,\psi)] := [(M \times N, \varphi \times \psi)]. \end{aligned}$$

I will leave it as an exercise to convince yourself that this operation is well defined, and to clarify precisely what it means to say that it is *natural*: in particular, for each fixed pair of integers $m, n \geqslant 0$, this product defines a natural transformation between two functors from the *product category* $\mathsf{Top} \times \mathsf{Top}$ to Ab.

27.8. Exercises.

Exercise 27.1. Prove Proposition 27.9 (isomorphisms have unique inverses).

Exercise 27.2. Verify the claim in Example 27.18 that $\mathrm{Hom} : \mathsf{Ab} \times \mathsf{Ab} \to \mathsf{Ab}$ defines a contravariant functor in its first variable and a covariant functor in its second variable.

Exercise 27.3. Suppose $\mathscr{A}$ is a category whose objects form a set X, such that for each pair $x, y \in X$, the set of morphisms $\mathrm{Hom}(x,y)$ contains either exactly one element or none. We can turn this into a binary relation by writing $x \bowtie y$ for every pair such that $\mathrm{Hom}(x,y) \neq \varnothing$.

(a) What properties does the relation $\bowtie$ need to have in order for it to define a category in the way indicated above?

(b) If $\mathscr{B}$ is another category whose objects form a set Y with morphisms determined by a binary relation $\bowtie$ as indicated above, what properties does a map $f : X \to Y$ need to have in order for it to define a functor from $\mathscr{A}$ to $\mathscr{B}$?

Exercise 27.4. In any category $\mathscr{C}$, each object X has an **automorphism group** (also called **isotropy** group) $\operatorname{Aut}(X)$, consisting of all the isomorphisms in $\operatorname{Hom}(X, X)$. A **groupoid** is a category in which all morphisms are also isomorphisms.

(a) Show that if $\mathscr{G}$ is a groupoid and Grp denotes the usual category of groups with homomorphisms, there exists a contravariant functor from $\mathscr{G}$ to Grp that assigns to each object X of $\mathscr{G}$ its automorphism group $\operatorname{Aut}(X)$. How does this functor act on morphisms $X \to Y$? Could you alternatively define it as a *covariant* functor? Conclude either way that whenever X and Y are isomorphic objects in $\mathscr{G}$ (meaning there exists an isomorphism in $\operatorname{Hom}(X, Y)$), the groups $\operatorname{Aut}(X)$ and $\operatorname{Aut}(Y)$ are isomorphic.

(b) Given a topological space X and two points x, y, let $\operatorname{Hom}(x, y)$ denote the set of homotopy classes (with fixed end points) of paths $[0, 1] \to X$ from x to y, and define a composition function $\operatorname{Hom}(x, y) \times \operatorname{Hom}(y, z) \to \operatorname{Hom}(x, z) : (\alpha, \beta) \mapsto \alpha \cdot \beta$ by the usual notion of concatenation of paths. Show that this notion of morphisms defines a groupoid whose objects are the points in X.[47] In this case, what are the automorphism groups $\operatorname{Aut}(x)$ and the isomorphisms $\operatorname{Aut}(y) \to \operatorname{Aut}(x)$ given by the functor in part (a)?

Exercise 27.5. Consider the category $\mathbb{K}$-Vect of vector spaces over a fixed field $\mathbb{K}$.

(a) Show that there is a covariant functor $\boldsymbol{\Delta}^2$ from $\mathbb{K}$-Vect to itself, assigning to each $V \in \mathbb{K}$-Vect the dual of its dual space $(V^*)^*$. Describe how this functor acts on morphisms.

(b) Construct a natural transformation from the identity functor $\mathsf{Id} : \mathbb{K}$-$\mathsf{Vect} \to \mathbb{K}$-$\mathsf{Vect}$ to $\boldsymbol{\Delta}^2$ that assigns to every $V \in \mathbb{K}$-Vect a linear injection $V \to (V^*)^*$, which is an isomorphism whenever V is finite dimensional.

Exercise 27.6. The **conjugate** $\bar{V}$ of a complex vector space V is defined as the same set $\bar{V} := V$ with the same notion of vector addition but with multiplication by scalars $\lambda = a + ib \in \mathbb{C}$ defined as multiplication by the complex conjugate $\bar{\lambda} = a - ib$. In other words, if $V \to \bar{V} : v \mapsto \bar{v}$ denotes the identity map, then scalar multiplication on $\bar{V}$ is defined so as to make this map complex antilinear, giving the formula

$$\lambda\bar{v} := \overline{\bar{\lambda}v} \in \bar{V} \qquad \text{for } \lambda \in \mathbb{C},\ v \in V.$$

(a) Show that there is a covariant functor $\boldsymbol{\kappa} : \mathbb{C}$-$\mathsf{Vect} \to \mathbb{C}$-$\mathsf{Vect}$ that sends each $V \in \mathbb{C}$-Vect to its conjugate $\bar{V}$, and describe how this functor acts on morphisms.

(b) Show that if T is a natural transformation from $\mathsf{Id} : \mathbb{C}$-$\mathsf{Vect} \to \mathbb{C}$-$\mathsf{Vect}$ to $\boldsymbol{\kappa} : \mathbb{C}$-$\mathsf{Vect} \to \mathbb{C}$-$\mathsf{Vect}$, then T assigns to each $V \in \mathbb{C}$-Vect the zero map $V \to \bar{V}$.
Hint: What does the naturality of T imply about the specific morphism $V \to V : v \mapsto iv$?

Comment: The map $V \to \bar{V} : v \mapsto \bar{v}$ is always a real-linear isomorphism, but it is not complex linear and is thus not a morphism in $\mathbb{C}$-Vect. Every finite-dimensional complex vector space is of course complex-linearly isomorphic to its conjugate, simply because both spaces have the same dimension, but the lack of any nontrivial natural transformation $\mathsf{Id} \to \boldsymbol{\kappa}$ is a symptom of the fact that there is generally no canonical way to define such isomorphisms.

Exercise 27.7. Prove Proposition 27.28 on the computation of $\Omega_0^{\mathrm{O}}(X)$ and $\Omega_0^{\mathrm{SO}}(X)$ for any space X.

---

[47] It is called the **fundamental groupoid** of X.

Exercise 27.8. Prove Lemma 27.33, showing that the natural map $h : \pi_1(X, p) \to \Omega_1^{\mathrm{SO}}(X)$ is a group homomorphism.
Hint: You are looking for an oriented bordism (Σ, Φ) *in which* Σ *is a compact surface with three boundary components—the simplest surface of this kind is a so-called "pair of pants," which has the topology of a disk with two holes cut out. Assuming* Σ *is a pair of pants, try to define* $\Phi : \Sigma \to X$ *by first thinking about which subset of* Σ *should be mapped to the base point of* X*. If you know anything about Morse theory, there is a relatively simple Morse-theoretic picture that will almost immediately lead to the construction you need: it involves the gradient flow of a Morse function* $f : \Sigma \to \mathbb{R}$ *that is constant on each boundary component and has exactly one critical point of index* 1 *in the interior.*

28. Axioms for homology theories

We will not yet define any specific homology theory in this lecture, but we shall introduce the standard set of axioms satisfied by homology theories, and demonstrate their usefulness in computations. Along the way, we encounter a fundamental tool from homological algebra: exact sequences.

28.1. The category of R-modules. The bordism theories Ω_n^{O} and Ω_n^{SO} in the previous lecture were defined as functors from Top to the category Ab of abelian groups, though we saw that the groups $\Omega_n^{\mathrm{O}}(X)$ can also be regarded as vector spaces over $\mathbb{Z}_2$. For homology theory, it is also possible to work entirely in the category Ab, but it is sometimes profitable to generalize this to a category that includes both abelian groups and vector spaces as special cases. This generalization does not require any extra effort, so we might as well work in the more general setting from the beginning.

Notation. For the rest of this course, unless otherwise noted, the symbol

$$R$$

will always denote a fixed commutative ring with unit, the choice of which will often not matter. We then denote by

$$R\text{-}\mathsf{Mod}$$

the **category of modules over** R, whose morphisms are the R-module homomorphisms. For two modules $G, H \in R\text{-}\mathsf{Mod}$, we will denote the set of R-module homomorphisms $G \to H$ (which is also an R-module) by

$$\mathrm{Hom}_R(G, H) := \mathrm{Hom}_{R\text{-}\mathsf{Mod}}(G, H)$$

whenever there is a need to specify R, but the abbreviated notation

$$\mathrm{Hom}(G, H) := \mathrm{Hom}_R(G, H)$$

can also be used when the context is clear. Similarly, we can denote the tensor product of two R-modules by $G \otimes_R H$ whenever R needs to be specified, but we will otherwise abbreviate it as

$$G \otimes H := G \otimes_R H.$$

A trivial R-module[48] is denoted by

$$0 \in R\text{-}\mathsf{Mod}.$$

For our purposes, abelian groups will be the most important special case of R-modules (see Example 28.1 below), and for that reason, we will sometimes abuse terminology and use the word "group" in places where the word "module" would be more appropriate.

---

[48] As with one-point spaces, there is not a unique trivial R-module, but there is a unique R-module isomorphism between any two of them.

EXAMPLE 28.1. All abelian groups $G \in \mathsf{Ab}$ can equivalently be regarded as modules over the commutative ring $\mathbb{Z}$, with scalar multiplication nx for $n \in \mathbb{Z}$ and $x \in G$ determined in the obvious way by the addition operation. Group homomorphisms are then automatically also $\mathbb{Z}$-module homomorphisms, and in this sense, the categories Ab and $\mathbb{Z}$-Mod are completely equivalent.

EXAMPLE 28.2. If R is a field $\mathbb{K}$, then an R-module is the same thing as a vector space over $\mathbb{K}$, and R-Mod is in this case equivalent to the category $\mathbb{K}$-Vect of vector spaces.

In this course, we will in practice almost exclusively be interested in the special cases where R is either $\mathbb{Z}$ or a field (most often either $\mathbb{Z}_2$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{K}$), and the category of R-modules will thus serve mainly as a single umbrella that encompasses both abelian groups and vector spaces.

One subtlety worth noting is that for any choice of the ring R, an R-module can always also be regarded as an abelian group, just by forgetting its scalar multiplication while keeping the addition operation, but doing this changes the definitions of tensor products $G \otimes H$ and the set of homomorphisms $\operatorname{Hom}(G, H)$. For instance, if $G, H \in \mathbb{R}$-Vect are real vector spaces, then they are also abelian groups and thus $\mathbb{Z}$-modules $G, H \in \mathbb{Z}$-Mod, but their tensor product in the sense of real vector spaces satisfies the relation

$$rg \otimes h = g \otimes rh \in G \otimes_{\mathbb{R}} H \qquad \text{for all} \qquad g \in G,\ h \in H,\ r \in \mathbb{R},$$

whereas the tensor product $G \otimes_{\mathbb{Z}} H$ in the sense of abelian groups only satisfies this when $r \in \mathbb{Z}$. Similarly, every $\mathbb{R}$-linear map $G \to H$ is also a homomorphism of abelian groups, but the converse is quite false.

DEFINITION 28.3. A **basis** of an R-module G is a subset $\mathcal{B} \subset G$ such that every element $g \in G$ can be written in the form

$$g = \sum_{b \in \mathcal{B}} g_b b$$

for some coefficients $g_b \in G$ that are uniquely determined by g, at most finitely-many of which are nonzero. An R-module is called **free** if it admits a basis.

A choice of basis $\mathcal{B} \subset G$ for a free R-module is equivalent to a choice of R-module isomorphism

$$\bigoplus_{b \in \mathcal{B}} R \xrightarrow{\cong} G,$$

so for instance, an abelian group (i.e. $\mathbb{Z}$-module) is free if and only if it is isomorphic to a direct sum of copies of $\mathbb{Z}$. Obviously, not every abelian group G has this property, e.g. it is never true if G is finite. On the other hand, a standard argument in linear algebra (using Zorn's lemma for the infinite-dimensional case) shows that every vector space admits a basis, so when R is a field, all R-modules are free. This basic fact is one of the key advantages of having the freedom to work with vector spaces instead of just abelian groups.

28.2. Exact sequences and splittings. In homological algebra, exact sequences play a role comparable to that of Cauchy sequences in analysis; that is to say, the entire subject would be impossible without them.

By a **sequence** (*Sequenz*) of R-modules, we mean a linearly ordered collection of modules A_n for $n \in \mathbb{Z}$, together with R-modules homomorphisms $\alpha_n : A_n \to A_{n+1}$. Depending on the context in which sequences arise, we can allow n to vary over any contiguous subset of the integers, which may be unbounded, or bounded above and/or below, so the sequence itself may have finitely or infinitely many terms, with or without a starting or end point. Let us call A_n an **interior** term of the sequence if the sequence also includes both A_{n-1} and A_{n+1}, thus giving rise to a three-term subsequence

$$A_{n-1} \xrightarrow{\alpha_{n-1}} A_n \xrightarrow{\alpha_n} A_{n+1}.$$

In this situation, we say that the sequence is **exact** (*exakt*) at the term A_n if

$$\operatorname{im}\alpha_{n-1} = \ker\alpha_n.$$

We do not define exactness for non-interior terms, i.e. terms that are at the beginning or end of the sequence. An **exact sequence** (*exakte Sequenz*) of R-modules is a sequence that is exact at all of its interior terms.

EXAMPLE 28.4. A sequence of the form $0 \longrightarrow A \xrightarrow{f} B \longrightarrow 0$ is exact if and only if f is an isomorphism.

EXAMPLE 28.5. An exact sequence with five terms that begins and ends with trivial modules

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is called a **short exact sequence** (*kurze exakte Sequenz*). Exactness means in this case that f is injective, g is surjective, and $\operatorname{im} f = \ker g$. A popular class of examples is the sequence

$$0 \to A \hookrightarrow B \xrightarrow{q} B/A \to 0$$

for any submodule $A \subset B$, where q denotes the quotient projection. Another is

$$0 \to A \overset{i}{\hookrightarrow} A \oplus C \xrightarrow{p} C \to 0 \tag{28.1}$$

for any two modules A and C, with the obvious inclusion map $i(a) := (a, 0)$ and projection map $p(a, c) := c$.

DEFINITION 28.6. A short exact sequence $0 \to A \to B \to C \to 0$ is said to **split**, and is then called a **split exact sequence**, if there exists an isomorphism $B \cong A \oplus C$ identifying it with the sequence in (28.1).

In the category of abelian groups, there are easy examples of short exact sequences that do not split, e.g. writing $q : \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} =: \mathbb{Z}_2$ for the quotient projection,

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\cdot 2} \mathbb{Z} \xrightarrow{q} \mathbb{Z}_2 \longrightarrow 0$$

is such an example, since $\mathbb{Z}$ is not isomorphic to $\mathbb{Z} \oplus \mathbb{Z}_2$. The next result, whose proof is a straightforward exercise, gives a useful practical criterion for short exact sequences to split, and its corollary implies in particular that they *always* split if R is a field.

THEOREM 28.7. *The following conditions on a short exact sequence* $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ *are equivalent:*

(i) The sequence splits;
(ii) The injective homomorphism $f : A \to B$ *admits a left-inverse* $B \to A$*;*
(iii) The surjective homomorphism $g : B \to C$ *admits a right-inverse* $C \to B$*.*

□

COROLLARY 28.8. *If C is a free R-module, then every short exact sequence* $0 \to A \to B \to C \to 0$ *splits.*

PROOF. Use a basis of C to define a right-inverse for the surjective map $B \to C$. □

Here is another popular application of exactness whose proof is an easy exercise.

THEOREM 28.9. *For an exact sequence of the form*

$$\ldots \longrightarrow A_n \xrightarrow{f_n} B_n \to C_n \to A_{n+1} \xrightarrow{f_{n+1}} B_{n+1} \longrightarrow C_{n+1} \longrightarrow \ldots,$$

the following conditions are equivalent:

(i) The modules C_n are trivial for every n;
(ii) The maps $f_n : A_n \to B_n$ are isomorphisms for every n.

□

28.3. Relative bordism groups. For a first real-life example of an exact sequence that arises naturally in topology, we can generalize the previous lecture's discussion and define **relative bordism groups**

$$\Omega_n^{\mathrm{O}}(X, A)$$

for every so-called **pair of spaces** (X, A), meaning a space X together with a choice of subset $A \subset X$. Given two pairs of spaces (X, A) and (Y, B), a **map of pairs**

$$f : (X, A) \to (Y, B) \qquad \text{or} \qquad (X, A) \xrightarrow{f} (Y, B)$$

is a continuous map $f : X \to Y$ such that $f(A) \subset B$, thus if we assign subspace topologies to A and B, the restriction $f|_A$ becomes a continuous map $A \to B$. Let us focus the discussion for now on *unoriented* bordism theory; the oriented case is completely analogous. Elements of $\Omega_n^{\mathrm{O}}(X, A)$ are equivalence classes $[(M, \varphi)]$ in which M is a compact smooth n-manifold that is allowed to have nonempty boundary, and φ is a map of pairs

$$(M, \partial M) \xrightarrow{\varphi} (X, A).$$

Two such pairs (M, φ) and (N, ψ) are equivalent if there is a **relative bordism** between them: this means a pair (W, Φ) consisting of a compact smooth $(n + 1)$-manifold W equipped with a smooth embedding

$$M \amalg N \hookrightarrow \partial W,$$

and a map of pairs

$$(W, \partial W \backslash (M \amalg N)) \xrightarrow{\Phi} (X, A)$$

such that $\Phi|_{M \amalg N} = \varphi \amalg \psi$. Note that while the domain of $\varphi : M \to X$ in this definition is allowed to have nonempty boundary, it may also be closed, thus the definition still makes sense if $A = \varnothing$ and just reproduces the so-called **absolute** bordism groups defined in the previous lecture,

$$\Omega_n^{\mathrm{O}}(X, \varnothing) = \Omega_n^{\mathrm{O}}(X).$$

The group structure of $\Omega_n^{\mathrm{O}}(X, A)$ is again defined via disjoint unions, and there is a straightforward way of associating to each map of pairs $f : (X, A) \to (Y, B)$ a group homomorphism

$$\Omega_n^{\mathrm{O}}(X, A) \xrightarrow{f_*} \Omega_n^{\mathrm{O}}(Y, B),$$

so that Ω_n^{O} becomes a functor

$$\mathsf{Top}^{\mathrm{rel}} \xrightarrow{\Omega_n^{\mathrm{O}}} \mathsf{Ab},$$

defined on the category $\mathsf{Top}^{\mathrm{rel}}$ of pairs of spaces, whose morphisms are maps of pairs. We can identify Top with the subcategory of $\mathsf{Top}^{\mathrm{rel}}$ whose objects are pairs of the form $(X, \varnothing)$, and then interpret $\Omega_n^{\mathrm{O}} : \mathsf{Top}^{\mathrm{rel}} \to \mathsf{Ab}$ as an extension of the previously-defined functor $\Omega_n^{\mathrm{O}} : \mathsf{Top} \to \mathsf{Ab}$.

For any pair (X, A) and $n \geqslant 1$, there is also a group homomorphism

$$\begin{aligned} \Omega_n^{\mathrm{O}}(X, A) &\xrightarrow{\partial_*} \Omega_{n-1}^{\mathrm{O}}(A), \\ [(M, \varphi)] &\longmapsto [(\partial M, \varphi|_{\partial M})], \end{aligned}$$

which is well defined because if (W, Φ) is a relative bordism between two representatives (M, φ) and (N, ψ), then restricting Φ to the compact n-manifold obtained by removing the interiors of M and N from ∂W defines an absolute bordism between $(\partial M, \varphi|_{\partial M})$ and $(\partial N, \psi|_{\partial N})$. One can interpret ∂_* as a natural transformation between two functors $\mathsf{Top}^{\mathrm{rel}} \to \mathsf{Ab}$, the details of which I will leave to the reader. What I really want to point out about ∂_* is the following:

Theorem 28.10. *Given a pair of spaces* (X, A)*, let* $i : A \hookrightarrow X$ *and* $j : (X, \varnothing) \hookrightarrow (X, A)$ *denote the obvious inclusions. Then the sequence of abelian groups*

$$\ldots \longrightarrow \Omega_n^{\mathrm{O}}(A) \xrightarrow{i_*} \Omega_n^{\mathrm{O}}(X) \xrightarrow{j_*} \Omega_n^{\mathrm{O}}(X, A) \xrightarrow{\partial_*} \Omega_{n-1}^{\mathrm{O}}(A) \xrightarrow{i_*} \Omega_{n-1}^{\mathrm{O}}(X) \longrightarrow \ldots$$

$$\ldots \longrightarrow \Omega_1^{\mathrm{O}}(X, A) \xrightarrow{\partial_*} \Omega_0^{\mathrm{O}}(A) \xrightarrow{i_*} \Omega_0^{\mathrm{O}}(X) \xrightarrow{j_*} \Omega_0^{\mathrm{O}}(X, A) \longrightarrow 0$$

is exact.

Corollary 28.11 (via Theorem 28.9). *For a pair of spaces* (X, A)*, the map* $\Omega_n^{\mathrm{O}}(A) \to \Omega_n^{\mathrm{O}}(X)$ *induced by the inclusion* $A \hookrightarrow X$ *is an isomorphism for every* $n \geqslant 0$ *if and only if* $\Omega_n^{\mathrm{O}}(X, A) = 0$ *for every* $n \geqslant 0$*.* ☐

We will later see an analogue of Theorem 28.10 in singular homology that plays a major role in that theory, and whose proof requires some elementary but non-obvious ideas from homological algebra. It's worth noting that the proof of Theorem 28.10, by comparison, is much more direct and straightforward; see Exercise 28.2.

28.4. The Eilenberg-Steenrod axioms. In the early history of homology, multiple packages of invariants were proposed that were easier to compute than the bordism groups, while seeming to measure similar topological information. The resulting theories differ in the details of their definitions—some of them drastically—but turn out to be naturally isomorphic if one restricts them to a "nice" class of spaces, which in practice includes all of the spaces that one is typically interested in, such as manifolds. Eventually, singular homology settled into a special role as the "standard" homology theory that everyone needs to learn, but in fact, one usually doesn't need to know its precise definition in order to use it. What's much more important are the *formal* properties that it satisfies, which are common to all homology theories, and were codified in the middle of the 20th century as a set of axioms due to Eilenberg and Steenrod [**ES52**], with a bit of extra input from Milnor [**Mil62**].

Definition 28.12. Fix as usual a commutative ring R with unit. An **axiomatic homology theory** h_* valued in the category of R-modules is a collection $\{h_n\}_{n \in \mathbb{Z}}$ of covariant functors

$$\mathsf{Top}^{\mathrm{rel}} \xrightarrow{h_n} R\text{-}\mathsf{Mod} : (X, A) \mapsto h_n(X, A)$$

defined for each $n \in \mathbb{Z}$, which also determine functors $h_n : \mathsf{Top} \to R\text{-}\mathsf{Mod}$ by defining

$$h_n(X) := h_n(X, \varnothing).$$

For a map of pairs $f : (X, A) \to (Y, B)$, the R-module homomorphism induced by the functor h_n is denoted by

$$h_n(X, A) \xrightarrow{f_*} h_n(Y, B).$$

The data of a homology theory also includes natural transformations ∂_* from the functor $\mathsf{Top}^{\mathrm{rel}} \to R\text{-}\mathsf{Mod} : (X, A) \mapsto h_n(X, A)$ to the functor $\mathsf{Top}^{\mathrm{rel}} \to R\text{-}\mathsf{Mod} : (X, A) \mapsto h_{n-1}(A)$ for each $n \in \mathbb{Z}$, and we require the following axioms:

- (Homotopy) For any two homotopic maps of pairs $f, g : (X, A) \to (Y, B)$, the induced morphisms $f_*, g_* : h_n(X, A) \to h_n(Y, B)$ are identical. (See Remark 28.14 below for the notion of a *homotopy of maps of pairs*.)

- (EXACTNESS) For all pairs (X, A) with inclusion maps $i : A \hookrightarrow X$ and $j : (X, \varnothing) \hookrightarrow (X, A)$, the sequence
$$\ldots \longrightarrow h_{n+1}(X, A) \xrightarrow{\partial_*} h_n(A) \xrightarrow{i_*} h_n(X) \xrightarrow{j_*} h_n(X, A) \xrightarrow{\partial_*} h_{n-1}(A) \longrightarrow \ldots$$
is exact.
- (EXCISION) For any pair (X, A) and any subset $B \subset A$ such that there exists a continuous function $u : X \to I$ equal to 0 on B and 1 on $X \backslash A$, the map induced by the inclusion $(X \backslash B, A \backslash B) \hookrightarrow (X, A)$ is an isomorphism
$$h_n(X \backslash B, A \backslash B) \xrightarrow{\cong} h_n(X, A).$$
- (DIMENSION) For any one-point space $\{*\}$, $h_n(\{*\}) = 0$ for all $n \neq 0$. The group $h_0(\{*\})$ is then called the **coefficient group** of the homology theory.[49]
- (ADDITIVITY) For any collection of spaces $\{X_\alpha\}_{\alpha \in J}$ with inclusion maps $i^\alpha : X_\alpha \hookrightarrow \coprod_{\beta \in J} X_\beta$, the map determined by the induced homomorphisms
$$i^\alpha_* : h_n(X_\alpha) \to h_n\left(\coprod_{\beta \in J} X_\beta\right)$$
is an isomorphism
$$\bigoplus_{\alpha \in J} i^\alpha_* : \bigoplus_{\alpha \in J} h_n(X_\alpha) \xrightarrow{\cong} h_n\left(\coprod_{\beta \in J} X_\beta\right).$$

You should be able to convince yourself without much trouble that the bordism functors $\Omega_n^{\mathrm{O}} : \mathsf{Top}^{\mathrm{rel}} \to \mathsf{Ab} = \mathbb{Z}\text{-}\mathsf{Mod}$ and their oriented counterparts Ω_n^{SO} each satisfy four out of the five Eilenberg-Steenrod axioms; see in particular Exercises 28.2 and 28.3. They do *not* satisfy the dimension axiom: this follows from Proposition 27.30 in the case of unoriented bordism theory, and there is a similar result for the oriented theory involving complex (instead of real) projective spaces. We call h_* a **generalized homology theory** if it satisfies all of the Eilenberg-Steenrod axioms except for dimension. In some contexts, the word "generalized" is removed, so that homology theories are typically assumed to satisfy four axioms instead of five, and those which also satisfy the dimension axiom are called **ordinary** homology theories. We will generally assume the dimension axiom in this semester and will not make use of any theories that don't satisfy it, but some of the results we prove about homology theories will be equally valid for generalized theories, since they do not depend on the dimension axiom.

A few further comments on the axioms are in order.

REMARK 28.13. The original list in [ES52] included three additional axioms at the beginning of the list, but the first two of these are equivalent to the statement that the h_n are functors, and the third simply requires ∂_* to be a natural transformation.

REMARK 28.14. The following definition is hopefully intuitive: a **homotopy** between two maps of pairs $f, g : (X, A) \to (Y, B)$ is a homotopy $H : X \times I \to Y$ between f and g such that $H(\cdot, t)$ is also a map of pairs $(X, A) \to (Y, B)$ for every $t \in I$, so in other words, H satisfies the condition
$$H(A \times I) \subset B.$$

---

[49] There is a slightly awkward semantic issue in this definition: strictly speaking, what we are calling "$\{*\}$" is not a unique space, but simply *any* choice of space that happens to contain only one element. It follows that the coefficient group $h_0(\{*\})$ is not a uniquely defined group, but is an isomorphism class of groups. Any two choices of one-point spaces P_0 and P_1 are related by a unique homeomorphism $P_0 \to P_1$, which induces a canonical isomorphism $h_0(P_0) \to h_0(P_1)$, and the coefficient group of a homology theory is unique in this sense.

We could also have chosen to hide the homotopy axiom by calling the h_n functors

$$\mathsf{hTop}^{\mathrm{rel}} \xrightarrow{h_n} \mathsf{Ab}_{\mathbb{Z}}$$

instead of $\mathsf{Top}^{\mathrm{rel}} \to \mathsf{Ab}_{\mathbb{Z}}$, where $\mathsf{hTop}^{\mathrm{rel}}$ denotes the **homotopy category of pairs of spaces**, having the same objects as $\mathsf{Top}^{\mathrm{rel}}$, but with homotopy classes of maps of pairs as morphisms. Note that a homotopy of maps of pairs $(X, \varnothing) \to (Y, \varnothing)$ is just a homotopy of maps $X \to Y$, making hTop naturally a subcategory of $\mathsf{hTop}^{\mathrm{rel}}$.

REMARK 28.15. The additivity axiom did not appear in [**ES52**], but was added later by Milnor [**Mil62**]. One can show in fact that for *finite* disjoint unions, additivity follows as a consequence of the other axioms (see Exercise 28.4), thus Eilenberg and Steenrod did not need it, because they were mainly concerned with computations for compact polyhedra—compactness precludes infinite disjoint unions.

REMARK 28.16. One often sees the excision axiom stated under a weaker hypothesis on the sets $B \subset A \subset X$, namely that the closure of B is contained in the interior of A. You might find it a challenge to think up an example in which that hypothesis is satisfied but the one we stated is not, and I don't encourage you to try, because within the class of spaces that are typically considered interesting to study, the two are fully equivalent; moreover, in all interesting situations I'm aware of, it is as easy to verify the stronger hypothesis as the weaker one. Singular homology does satisfy excision under the weaker hypothesis, but the existence of a function $u : X \to I$ separating B from $X \backslash A$ is a more natural condition from other points of view, especially in homotopy-theoretic reformulations of homology. The hypotheses originally stated in [**ES52**] also required B to be open, which is another detail that makes no meaningful difference for the class of spaces typically of interest.

REMARK 28.17. The reason the dimension axiom has the name that it does is that if it were not included in the list of axioms, then for every homology theory h_*, one could use arbitrary degree shifts to define new homology theories such as k_* with $k_n(X, A) := h_{n+1}(X, A)$. The dimension axiom prevents this, in the hope that the value of the subscript n in $h_n(X, A)$ will then have some geometric meaning. The reason for calling $h_0(\{*\})$ a "coefficient group" will become clearer when we write down concrete examples of homology theories.

REMARK 28.18. It is sometimes useful to expand the definition and allow an axiomatic homology theory to be a functor $\mathscr{C} \to R\text{-}\mathsf{Mod}$ defined on a suitable subcategory $\mathscr{C}$ of $\mathsf{Top}^{\mathrm{rel}}$, so that we need not define $h_*(X, A)$ for *all* pairs (X, A), but only a subclass. One useful example is the category of **compact pairs**, which are simply pairs of spaces (X, A) such that X is compact Hausdorff and $A \subset X$ is closed. Others include the categories of *polyhedra* and *CW-complexes*, which we'll have more to say about in future lectures. When allowing restrictions of this type, one must take care so that all of the maps needed for expressing the axioms—e.g. the inclusions $A \hookrightarrow X$ and $(X, \varnothing) \hookrightarrow (X, A)$—are actually morphisms in the category $\mathscr{C}$. In [**ES52**], this concern motivates the definition of the notion of an *admissible* category of pairs, though we have no need to reproduce that definition here.

28.5. Reduced homology. Assume h_* is a collection of functors as in Definition 28.12 satisfying at least the homotopy and exactness axioms. For technical reasons that will become clearer in the next section, it is sometimes useful to replace the groups $h_n(X)$ with certain subgroups $\widetilde{h}_n(X) \subset h_n(X)$ called **reduced homology** groups. To define them, we denote by

$$X \xrightarrow{\epsilon} \{*\}$$

the unique map from any given space X to a one-point space. We then use the induced homomorphisms $\epsilon_* : h_n(X) \to h_n(\{*\})$ to define

$$\widetilde{h}_n(X) := \ker\left(h_n(X) \xrightarrow{\epsilon_*} h_n(\{*\})\right) \subset h_n(X).$$

Observe that if h_* satisfies the dimension axiom, then $\widetilde{h}_n(X) = h_n(X)$ for all $n \ne 0$. If $n = 0$ or the dimension axiom is not satisfied, then we can typically expect $\widetilde{h}_n(X)$ and $h_n(X)$ to be different, and the best way to relate them to each other is through a split exact sequence. Indeed, observe that the map $\epsilon : X \to \{*\}$ is not only trivially surjective, but also admits a right-inverse, defined by choosing any embedding

$$\{*\} \overset{i}{\hookrightarrow} X.$$

It then follows from functoriality that the homomorphism $\epsilon_* : h_n(X) \to h_n(\{*\})$ likewise is surjective and admits a right-inverse, thus by Theorem 28.7,

$$0 \to \widetilde{h}_n(X) \hookrightarrow h_n(X) \overset{\epsilon_*}{\to} h_n(\{*\}) \to 0$$

is a split exact sequence, implying the existence of an isomorphism

$$h_n(X) \cong \widetilde{h}_n(X) \oplus h_n(\{*\}).$$

If h_* satisfies the dimension axiom and has coefficient group $G = h_0(\{*\})$, this becomes

$$h_n(X) \cong \begin{cases} \widetilde{h}_n(X) & \text{if } n \ne 0, \\ \widetilde{h}_n(X) \oplus G & \text{if } n = 0. \end{cases}$$

One should keep in mind however that this isomorphism is not generally canonical: it depends on the choice of inclusion $i : \{*\} \hookrightarrow X$, which determines the splitting of the exact sequence relating $\widetilde{h}_n(X)$ and $h_n(X)$.

Let us clarify why $\widetilde{h}_n$ for each $n \in \mathbb{Z}$ is naturally also a functor $\mathsf{Top} \to R\text{-}\mathsf{Mod}$.

Proposition 28.19. *The homomorphisms $f_* : h_n(X) \to h_n(Y)$ induced by any continuous map $f : X \to Y$ send $\widetilde{h}_n(X)$ into $\widetilde{h}_n(Y)$.*

Proof. Denote $\epsilon^X : X \to \{*\}$ and $\epsilon^Y : Y \to \{*\}$ for the unique maps, and notice that $\epsilon^Y \circ f = \mathrm{Id} \circ \epsilon^X$, thus the following diagram commutes.

$$\begin{tikzcd}
h_n(X) \arrow[r, "f_*"] \arrow[d, "\epsilon^X_*"] & h_n(Y) \arrow[d, "\epsilon^Y_*"] \\
h_n(\{*\}) \arrow[r, "\mathbb{1}"] & h_n(\{*\})
\end{tikzcd}$$

This implies that $f_*(\ker \epsilon^X_*) \subset \ker \epsilon^Y_*$. □

The next result reveals the main advantage of using $\widetilde{h}_*$ in place of h_* for certain applications.

Proposition 28.20. *If X is a contractible space, then $\widetilde{h}_n(X) = 0$ for every n.*

Proof. Contractibility implies that the map $\epsilon : X \to \{*\}$ is a homotopy equivalence, thus by the homotopy axiom, $\epsilon_* : h_n(X) \to h_n(\{*\})$ is an isomorphism, and its kernel $\widetilde{h}_n(X)$ is therefore trivial. □

Remark 28.21. If h_* also satisfies the dimension axiom, then we also have $h_n(X) = 0$ for all $n \ne 0$ whenever X is contractible, but $h_0(X)$ is typically nontrivial, as it is isomorphic to the coefficient group. As a consequence, some of the standard applications of reduced homology can also be carried out with unreduced homology, but only if the degree 0 groups are excluded from consideration.

The relative version of reduced homology is defined in a trivial way: we set

$$\widetilde{h}_n(X,A) := h_n(X,A) \quad \text{whenever} \quad A \neq \varnothing.$$

This seemingly naive definition is justified by the following considerations. Note first that the functors $\widetilde{h}_n : \mathsf{Top} \to R\text{-}\mathsf{Mod}$ now extend to pairs as functors $\mathsf{Top}^{\mathrm{rel}} \to R\text{-}\mathsf{Mod}$; here there is nothing to check since the existence of a map of pairs $(X,A) \to (Y,B)$ with $A \neq \varnothing$ implies $B \neq \varnothing$, so that both reduced relative homology groups match the unreduced case. Next, observe that for any space X, the relative homology groups $h_n(X,X)$ all vanish; this follows from the exactness axiom and Theorem 28.9, as we have an exact sequence

$$\ldots \longrightarrow h_n(X) \xrightarrow{\mathbb{1}} h_n(X) \longrightarrow h_n(X,X) \xrightarrow{\partial_*} h_{n-1}(X) \xrightarrow{\mathbb{1}} h_{n-1}(X) \longrightarrow \ldots$$

It follows that $\widetilde{h}_n(X,A)$ for $A \neq \varnothing$ is in fact the kernel of the map

$$h_n(X,A) \xrightarrow{\epsilon_*} h_*(\{*\},\{*\}) = 0$$

induced by the unique map of pairs $\epsilon : (X,A) \to (\{*\},\{*\})$. Moreover, the naturality of the connecting homomorphisms ∂_* gives a commutative diagram

$$\begin{tikzcd}
h_{n+1}(X,A) \arrow[r, "\partial_*"] \arrow[d, "\epsilon_*"] & h_n(A) \arrow[d, "\epsilon_*"] \\
h_{n+1}(\{*\},\{*\}) \arrow[r, "\partial_*"] & h_n(\{*\})
\end{tikzcd}$$

Since the term $h_{n+1}(\{*\},\{*\})$ is trivial, this diagram proves that the image of $\partial_* : h_{n+1}(X,A) \to h_n(A)$ is always in the subgroup $\widetilde{h}_n(A)$. We can therefore write down a well-defined sequence of homomorphisms

$$\ldots \longrightarrow \widetilde{h}_{n+1}(X,A) \xrightarrow{\partial_*} \widetilde{h}_n(A) \xrightarrow{i_*} \widetilde{h}_n(X) \xrightarrow{j_*} \widetilde{h}_n(X,A) \xrightarrow{\partial_*} \widetilde{h}_{n-1}(A) \longrightarrow \ldots$$

using the usual inclusions $i : A \hookrightarrow X$ and $j : (X,\varnothing) \hookrightarrow (X,A)$. It is not immediately obvious whether this sequence is exact, but consider the commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\ldots \arrow[r] & \widetilde{h}_n(A) \arrow[r, "i_*"] \arrow[d, hook] & \widetilde{h}_n(X) \arrow[r, "j_*"] \arrow[d, hook] & \widetilde{h}_n(X,A) \arrow[r, "\partial_*"] \arrow[d, equal] & \widetilde{h}_{n-1}(A) \arrow[r] \arrow[d, hook] & \ldots \\
\ldots \arrow[r] & h_n(A) \arrow[r, "i_*"] \arrow[d, "\epsilon_*"] & h_n(X) \arrow[r, "j_*"] \arrow[d, "\epsilon_*"] & h_n(X,A) \arrow[r, "\partial_*"] \arrow[d, "\epsilon_*"] & h_{n-1}(A) \arrow[r] \arrow[d, "\epsilon_*"] & \ldots \\
\ldots \arrow[r] & h_n(\{*\}) \arrow[r, equal] \arrow[d] & h_n(\{*\}) \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & h_{n-1}(\{*\}) \arrow[r] \arrow[d] & \ldots \\
 & 0 & 0 & 0 & 0 &
\end{tikzcd}$$

Here the bottom two nontrivial rows are exact due to the exactness axiom, and all columns in the diagram are short exact sequences by construction. The rest is algebra:

PROPOSITION 28.22. *Assume the following diagram of R-modules commutes, all its columns are exact sequences, and the bottom two nontrivial rows are also exact sequences:*

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\ldots & A_{n+2} \arrow[d, "\iota_{n+2}"] & A_{n+1} \arrow[d, "\iota_{n+1}"] & A_n \arrow[d, "\iota_n"] & A_{n-1} \arrow[d, "\iota_{n-1}"] & A_{n-2} \arrow[d, "\iota_{n-2}"] & \ldots \\
\ldots \arrow[r] & B_{n+2} \arrow[r, "g_{n+2}"] \arrow[d, "\epsilon_{n+2}"] & B_{n+1} \arrow[r, "g_{n+1}"] \arrow[d, "\epsilon_{n+1}"] & B_n \arrow[r, "g_n"] \arrow[d, "\epsilon_n"] & B_{n-1} \arrow[r, "g_{n-1}"] \arrow[d, "\epsilon_{n-1}"] & B_{n-2} \arrow[r] \arrow[d, "\epsilon_{n-2}"] & \ldots \\
\ldots \arrow[r] & C_{n+2} \arrow[r, "h_{n+2}"] \arrow[d] & C_{n+1} \arrow[r, "h_{n+1}"] \arrow[d] & C_n \arrow[r, "h_n"] \arrow[d] & C_{n-1} \arrow[r, "h_{n-1}"] \arrow[d] & C_{n-2} \arrow[r] \arrow[d] & \ldots \\
 & 0 & 0 & 0 & 0 & 0 &
\end{tikzcd}$$

Then the top nontrivial row can be endowed uniquely with maps $f_n : A_n \to A_{n-1}$ such that the diagram still commutes, and these make that row into an exact sequence.

PROOF. The method behind this proof is commonly known as *diagram chasing*, and we will later see several other examples of it. The basic idea is straightforward: at every step, we examine a particular term in the diagram, consider what is already known about the maps going into and out of that term, and then deduce whatever we can from given conditions such as exactness. In typical situations, whatever can be deduced tells you which term to examine in the next step.

If $f_n : A_n \to A_{n-1}$ can be defined so that the diagram commutes, then for $a \in A_n$ we need $f_n(a) \in \iota_{n-1}^{-1}(g_n\iota_n(a))$, and this condition will fully determine $f_n(a) \in A_{n-1}$ since ι_{n-1} is injective due to the exactness of columns. To see that the condition can be achieved, notice

$$\epsilon_{n-1}g_n\iota_n = h_n\epsilon_n\iota_n = 0,$$

thus $g_n\iota_n(a) \in \ker\epsilon_{n-1} = \operatorname{im}\iota_{n-1}$. This gives an element $x \in A_{n-1}$ such that $\iota_{n-1}(x) = g_n\iota_n(a)$, so we can set $f_n(a) = x$.

The goal is now to show that $\ldots A_{n+1} \overset{f_{n+1}}{\to} A_n \overset{f_n}{\to} A_{n-1} \to \ldots$ is an exact sequence. For each n, commutativity of the diagram gives

$$\iota_{n-2}f_{n-1}f_n = g_{n-1}g_n\iota_n = 0$$

since the middle row is exact, and the exactness of the columns implies in turn that ι_{n-2} is injective, thus $f_{n-1}f_n = 0$. To finish, we need to prove that every $a \in A_n$ satisfying $f_n(a) = 0$ also satisfies $a = f_{n+1}(x)$ for some $x \in A_{n+1}$. Using commutativity, we have

$$0 = \iota_{n-1}f_n(a) = g_n\iota_n(a),$$

thus the exactness of the middle row gives an element $b \in B_{n+1}$ such that $g_{n+1}(b) = \iota_n(a)$. If we knew $\epsilon_{n+1}(b) = 0$, then we could at this point appeal to the exactness of the columns and write $b = \iota_{n+1}(x)$ for some $x \in A_{n+1}$, which would then satisfy $\iota_n f_{n+1}(x) = g_{n+1}\iota_{n+1}(x) = g_{n+1}(b) = \iota_n(a)$ and therefore $f_{n+1}(x) = a$ since ι_n is injective. But $\epsilon_{n+1}(b)$ might not be 0, so to finish the proof, we claim instead that b can be replaced by another element $b' \in B_{n+1}$ that satisfies both $g_{n+1}(b') = \iota_n(a)$ and $\epsilon_{n+1}(b') = 0$.

To find b', observe that by commutativity and the exactness of the columns,

$$h_{n+1}\epsilon_{n+1}(b) = \epsilon_n g_{n+1}(b) = \epsilon_n\iota_n(a) = 0,$$

thus by the exactness of the bottom row, $\epsilon_{n+1}(b) = h_{n+2}(c)$ for some $c \in C_{n+2}$. Appealing again to the exactness of the columns, ϵ_{n+2} is surjective, so we have $c = \epsilon_{n+2}(y)$ for some $y \in B_{n+2}$. Set

$$b' := b - g_{n+2}(y).$$

This satisfies $g_{n+1}(b') = g_{n+1}(b) - g_{n+1}g_{n+2}(y) = g_{n+1}(b) = \iota_n(a)$, and using commutativitiy again,

$$\epsilon_{n+1}(b') = \epsilon_{n+1}(b) - \epsilon_{n+1}g_{n+2}(y) = \epsilon_{n+1}(b) - h_{n+2}\epsilon_{n+2}(y) = \epsilon_{n+1}(b) - h_{n+2}(c) = 0.$$

□

We have proved:

THEOREM 28.23. *For any pair of spaces (X, A) and any homology theory h_*, there is an exact sequence of reduced homology groups*

$$\ldots \longrightarrow \widetilde{h}_{n+1}(X, A) \xrightarrow{\partial_*} \widetilde{h}_n(A) \xrightarrow{i_*} \widetilde{h}_n(X) \xrightarrow{j_*} \widetilde{h}_n(X, A) \xrightarrow{\partial_*} \widetilde{h}_{n-1}(A) \longrightarrow \ldots,$$

where $i : A \hookrightarrow X$ and $j : (X, \varnothing) \hookrightarrow (X, A)$ are the obvious inclusions and $\partial_* : \widetilde{h}_n(X, A) \to \widetilde{h}_{n-1}(A)$ is the same map as the usual connecting homomorphism $h_n(X, A) \to h_{n-1}(A)$. □

28.6. Suspension isomorphisms. The following general construction leads easily to a complete computation of $h_*(S^n)$ for any axiomatic homology theory. We assume in this section that h_* is a generalized homology theory, so it satisfies all the conditions in Definition 28.12 except possibly the dimension axiom.[50]

Recall that for an arbitrary space X, the **suspension** (*Einhängung*) of X is a space ΣX formed by gluing together two cones $C_+X := CX := (X \times [0, 1])/(X \times \{1\})$ and $C_-X := (X \times [-1, 0])/(X \times \{-1\})$ along $X = X \times \{0\} \subset C_\pm X$, in short,

$$\Sigma X := C_+X \cup_X C_-X.$$

THEOREM 28.24. *For every space X, integer $k \in \mathbb{Z}$ and generalized homology theory h_*, the diagram (28.2) below gives rise to a natural isomorphism*

$$\Sigma_* := \varphi_*^{-1} \circ j_* \circ i_* \circ \partial_*^{-1} : \widetilde{h}_k(X) \to \widetilde{h}_{k+1}(\Sigma X).$$

PROOF. Let

$$p_+ \in C_+X \subset \Sigma X \qquad \text{and} \qquad p_- \in C_-X \subset \Sigma X$$

denote the summits of the two cones that are glued together to form the suspension, e.g. if we write $C_+X = (X \times [0, 1])/(X \times \{1\})$, then $p_+ \in C_+X$ is the point that results from collapsing $X \times \{1\}$. We then consider the diagram

(28.2)
$$\begin{tikzcd}
\widetilde{h}_k(X) & & \widetilde{h}_{k+1}(\Sigma X) \arrow[d, "\varphi_*"] \\
\widetilde{h}_{k+1}(C_+X, X) \arrow[u, "\partial_*"] \arrow[r, "i_*"] & \widetilde{h}_{k+1}(\Sigma X \backslash \{p_-\}, C_-X \backslash \{p_-\}) \arrow[r, "j_*"] & \widetilde{h}_{k+1}(\Sigma X, C_-X)
\end{tikzcd}$$

in which three of the maps are determined by the obvious inclusions of pairs,

$$(C_+X, X) \overset{i}{\hookrightarrow} (\Sigma X \backslash \{p_-\}, C_-X \backslash \{p_-\}),$$
$$(\Sigma X \backslash \{p_-\}, C_-X \backslash \{p_-\}) \overset{j}{\hookrightarrow} (\Sigma X, C_-X),$$
$$(\Sigma X, \varnothing) \overset{\varphi}{\hookrightarrow} (\Sigma X, C_-X).$$

---

[50] In fact, the additivity axiom is also not strictly necessary for this discussion, since by Exercise 28.4, it follows from the other axioms in the case of finite disjoint unions.

The first of these is a homotopy equivalence, as there exists a deformation retraction of the pair $(\Sigma X \backslash \{p_-\}, C_-X\backslash\{p_-\})$ to (C_+X, X), thus i_* is an isomorphism by the homotopy axiom. Since one can easily define a function $u : \Sigma X \to I$ that vanishes at p_- and equals 1 on C_+X, the excision axiom implies that j_* is also an isomorphism. For the other two maps, we consider the exact sequences provided by Theorem 28.23 for the pairs $(\Sigma X, C_-X)$ and (C_+X, X), that is

$$\ldots \longrightarrow \widetilde{h}_{k+1}(C_-X) \longrightarrow \widetilde{h}_{k+1}(\Sigma X) \xrightarrow{\varphi_*} \widetilde{h}_{k+1}(\Sigma X, C_-X) \longrightarrow \widetilde{h}_k(C_-X) \longrightarrow \ldots$$

and

$$\ldots \longrightarrow \widetilde{h}_{k+1}(C_+X) \longrightarrow \widetilde{h}_{k+1}(C_+X, X) \xrightarrow{\partial_*} \widetilde{h}_k(X) \longrightarrow \widetilde{h}_k(C_+X) \longrightarrow \ldots$$

The contractibility of $C_\pm X$ implies via Proposition 28.20 that

$$\widetilde{h}_k(C_\pm X) \cong \widetilde{h}_k(\{*\}) = 0, \qquad \text{and} \qquad \widetilde{h}_{k+1}(C_\pm X) \cong \widetilde{h}_{k+1}(\{*\}) = 0,$$

thus the exactness of these two sequences implies that φ_* and ∂_* are both isomorphisms.

The *naturality* of the map $\Sigma_* : \widetilde{h}_k(X) \to \widetilde{h}_{k+1}(\Sigma X)$ has a precise meaning, because the suspension operation can be understood as a functor $\Sigma : \mathsf{Top} \to \mathsf{Top}$, and the statement is then that Σ_* defines a natural transformation between two functors $\mathsf{Top} \to R\text{-}\mathsf{Mod}$, namely $\widetilde{h}_k$ and $\widetilde{h}_{k+1} \circ \Sigma$. This follows in a straightforward way using the naturality of the homomorphisms ∂_*; the details are an exercise. □

28.7. Homology groups of spheres. Recall that the suspension of a sphere is also a sphere, but one dimension higher:

$$\Sigma(S^n) \cong S^{n+1}.$$

This fact and Theorem 28.24 make possible an inductive computation of $h_*(S^n)$ for every axiomatic homology theory and every $n \geqslant 0$, using the fact that S^0 is the disjoint union of two one-point spaces. Here is the statement; the proof is Exercise 28.6.

THEOREM 28.25. *Assume h_* is an axiomatic homology theory with coefficient group $h_0(\{*\}) = G$. Then for each pair of integers $k \in \mathbb{Z}$ and $n \geqslant 1$,*

$$h_k(S^n) \cong \begin{cases} G & \textit{if } k = 0 \textit{ or } k = n, \\ 0 & \textit{otherwise.} \end{cases}$$

□

28.8. Exercises.

EXERCISE 28.1 (*). Prove Theorem 28.7 on split exact sequences, and Theorem 28.9 on long exact sequences with every third term vanishing.

EXERCISE 28.2. Prove that the sequence of relative and absolute bordism groups in Theorem 28.10 is exact. Here are a couple of hints:

- For exactness at $\Omega_n^{\mathrm{O}}(X)$: $j_*[(M, \varphi)] = 0$ means $\varphi : M \to X$ can be extended over a compact $(n+1)$-manifold W with $M \subset \partial W$ such that the extension maps $\partial W \backslash M$ into A. In this situation, M is a *closed* manifold—what does that imply about $\partial W \backslash M$?
- For exactness at $\Omega_n^{\mathrm{O}}(X, A)$: $\partial_*[(M, \varphi)] = 0$ means that $\varphi|_{\partial M} : \partial M \to A$ extends to a map $V \to A$ on some compact n-manifold V whose boundary is identified with ∂M. Build a closed n-manifold out of V and M.

EXERCISE 28.3. Prove that the bordism theories Ω_*^{O} and Ω_*^{SO} satisfy the homotopy, excision, and additivity axioms.
Hint for excision: Suppose $u : X \to I$ is a function as specified in the excision axiom, and $\varphi : (M, \partial M) \to (X, A)$ is a map of pairs so that (M, φ) represents a bordism class. By standard

results about smooth manifolds, the function $u \circ \varphi : M \to I$ *can be perturbed to a function* $v : M \to I$ *such that for some* $r \in (0,1)$*, both* $v^{-1}(r) \subset M$ *and* $(v|_{\partial M})^{-1}(r) \subset \partial M$ *are smooth submanifolds.*

Exercise 28.4. Assume h_* is a collection of functors $h_n : \mathsf{Top}^{\mathrm{rel}} \to R\text{-}\mathsf{Mod}$ for $n \in \mathbb{Z}$ satisfying the exactness and excision axioms of Eilenberg-Steenrod. Given two spaces X, Y and the natural inclusions $i^X : X \hookrightarrow X \amalg Y$ and $i^Y : Y \hookrightarrow X \amalg Y$, show that the map

$$i^X_* \oplus i^Y_* : h_n(X) \oplus h_n(Y) \to h_n(X \amalg Y) : (x, y) \mapsto i^X_* x + i^Y_* y$$

is an isomorphism, and deduce that h_* also satisfies the additivity axiom for all *finite* disjoint unions.
Hint: Apply exactness and excision to the pairs $(X \amalg Y, X)$ *and* $(X \amalg Y, Y)$.

Exercise 28.5 (*). Assume h_* is an axiomatic homology theory with coefficient group $h_0(\{*\}) = G$. For any two spaces X and Y with maps $\epsilon^X : X \to \{*\}$ and $\epsilon^Y : Y \to \{*\}$, show that the natural isomorphism $h_n(X \amalg Y) \cong h_n(X) \oplus h_n(Y)$ identifies $\widetilde{h}_n(X \amalg Y)$ with $\ker(\epsilon^X_* \oplus \epsilon^Y_*) \subset h_n(X; G) \oplus h_n(Y; G)$. Then apply this in the case $X = Y = \{*\}$ to identify $\widetilde{h}_0(\{*\} \amalg \{*\})$ with the kernel of the map

$$\mathbb{1} \oplus \mathbb{1} : G \oplus G \to G : (g, h) \mapsto g + h,$$

which is isomorphic to G.

Exercise 28.6 (*). Given an axiomatic homology theory h_* with coefficient group G, use Theorem 28.24, Exercise 28.5 and an inductive argument to derive a general formula for $\widetilde{h}_k(S^n)$ for all $k \in \mathbb{Z}$ and $n \geqslant 0$, and then deduce from it Theorem 28.25.

Exercise 28.7. One of the most popular simple applications of homology is the Brouwer fixed point theorem, which states that for the closed disk $\mathbb{D}^n \subset \mathbb{R}^n$ of any dimension $n \in \mathbb{N}$, every continuous map $f : \mathbb{D}^n \to \mathbb{D}^n$ has a fixed point.

(a) Deduce the Brouwer fixed point theorem from the following statement: For each $n \in \mathbb{N}$, the disk $\mathbb{D}^n$ does not admit any retraction to its boundary $\partial \mathbb{D}^n = S^{n-1}$.
Hint: If $f : \mathbb{D}^n \to \mathbb{D}^n$ *has no fixed points, then there is a unique line through* x *and* $f(x)$ *for every* $x \in \mathbb{D}^n$.

(b) Assuming the existence of an axiomatic homology theory h_* with a nontrivial coefficient group, deduce from the computation of $h_*(S^{n-1})$ that retractions $\mathbb{D}^n \to S^{n-1}$ cannot exist.

Exercise 28.8 (*). The subject of this exercise is a standard tool in homological algebra known as the **five-lemma**.

(a) Suppose the following diagram commutes and that both of its rows are exact, meaning $\operatorname{im} f = \ker g$, $\operatorname{im} g' = \ker h'$ and so forth:

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C \arrow[r, "h"] \arrow[d, "\gamma"] & D \arrow[r, "i"] \arrow[d, "\delta"] & E \arrow[d, "\varepsilon"] \\
A' \arrow[r, "f'"] & B' \arrow[r, "g'"] & C' \arrow[r, "h'"] & D' \arrow[r, "i'"] & E'
\end{tikzcd}$$

Prove that if α, β, δ and ε are all isomorphisms, then so is γ.

(b) Here is an application: given a homology theory h_* and a map of pairs $f : (X, A) \to (Y, B)$, show that if any two of the induced maps $f_* : h_n(X) \to h_n(Y)$, $f_* : h_n(A) \to h_n(B)$ and $f_* : h_n(X, A) \to h_n(Y, B)$ are isomorphisms for every n, then so is the third.

(c) Prove that every homology theory h_* also satisfies a relative version of the additivity axiom, involving disjoint unions of pairs of spaces

$$\coprod_{\beta \in J} (X_\beta, A_\beta) := \left(\coprod_{\beta \in J} X_\beta, \coprod_{\beta \in J} A_\beta \right).$$

29. Simplicial homology

As mental preparation for the definition of singular homology, it will be helpful to start with a different theory that is similar but more restrictive. Simplicial homology requires strictly more data for its definition than just a topological space, and thus can only be defined on spaces that are "nice" enough to admit such data. What it lacks in generality, it makes up for in computability and geometric transparency. One can think of simplicial homology as a combinatorial variant of bordism theory, one that is based on simpler building blocks than manifolds, and can thus be studied without any understanding of the (generally difficult) problem of classifying manifolds.

29.1. Simplicial complexes and polyhedra. The spaces on which simplicial homology is defined are called polyhedra, and they are much more restrictive than arbitrary topological spaces, but nonetheless include most of the typical examples of interest, e.g. all smooth manifolds. Intuitively, a polyhedron is a space that can be constructed by gluing together "triangles" of various dimensions, and the resulting decomposition of a polyhedron into "triangular" pieces is therefore known as a *triangulation.* The first necessary step is to define the n-dimensional generalization of a triangle.

DEFINITION 29.1. For an integer $n \geqslant 0$, the **standard n-simplex** is the topological space

$$\Delta^n := \left\{ (t_0, \ldots, t_n) \in I^{n+1} \;\middle|\; t_0 + \ldots + t_n = 1 \right\},$$

endowed with the subspace topology as a subset of $\mathbb{R}^{n+1}$. The $n+1$ standard basis vectors of $\mathbb{R}^{n+1}$ are called the **vertices** (*Eckpunkte*) of Δ^n, and for arbitrary subsets $J \subset \{0, \ldots, n\}$, the sets of the form

$$\left\{ (t_0, \ldots, t_n) \in \Delta^n \;\middle|\; t_j = 0 \text{ for all } j \in J \right\}$$

are called the **faces** (*Seiten* or *Facetten*) of Δ^n; these include in particular the $n+1$ **boundary faces** (*Seitenflächen*)

$$\partial_{(j)} \Delta^n := \left\{ (t_0, \ldots, t_n) \in \Delta^n \;\middle|\; t_j = 0 \right\}, \qquad j = 0, \ldots, n.$$

This definition makes Δ^0 the one-point space $\{1\} \subset \mathbb{R}$, while Δ^1 is a compact line segment in $\mathbb{R}^2$ homeomorphic to the interval I, Δ^2 is the compact region in a plane bounded by a triangle, Δ^3 is the compact region in a 3-dimensional vector space bounded by a tetrahedron, and so forth. Observe that every face of Δ^n is homeomorphic to Δ^k for some $k \leqslant n$, and since the coordinates of $\mathbb{R}^{n+1}$ come with a canonical ordering, there is even a canonical choice of homeomorphism. For instance, the boundary faces $\partial_{(j)} \Delta^n$ are all homeomorphic to Δ^{n-1}, and the canonical homeomorphisms take the form

$$\Delta^{n-1} \xrightarrow{\cong} \partial_{(j)} \Delta^n : (t_0, \ldots, t_{n-1}) \mapsto (t_0, \ldots, t_{j-1}, 0, t_j, \ldots, t_{n-1}). \tag{29.1}$$

We will make frequent use of these canonical homeomorphisms to identify each face of a standard simplex with another standard simplex.

In order to explain how copies of Δ^n for various $n \geqslant 0$ can be glued together to form a polyhedron, we need to define simplicial complexes, which are fundamentally combinatorial objects.

Definition 29.2. A **simplicial complex** (*Simplizialkomplex*) $K = (V, S)$ consists of two sets V and S, called the sets of **vertices** (*Eckpunkte*) and **simplices** (*Simplizes*) respectively, where the elements of S are finite subsets of V, and $\sigma \in S$ is called an **n-simplex** of K if it has $n + 1$ elements. We require the following conditions:

(1) Every vertex $v \in V$ gives rise to a 0-simplex in K, i.e. $\{v\} \in S$;
(2) If $\sigma \in S$ then every subset $\sigma' \subset \sigma$ is also an element of S.

For any n-simplex $\sigma \in S$, its subsets are called its **faces** (*Seiten* or *Facetten*), and in particular the subsets that are $(n-1)$-simplices are called **boundary faces** (*Seitenflächen*) of σ. The second condition above thus says that for every simplex in the complex, all of its faces also belong to the complex. With this condition in place, the first condition is then equivalent to the requirement that every vertex in the set V belongs to at least one simplex.

The complex K is said to be **finite** if V (and therefore also S) is finite, and its **dimension** is

$$\dim K := \sup_{\sigma \in S} \dim \sigma \in \{0, 1, 2, \ldots, \infty\},$$

where we write $\dim \sigma = n$ whenever σ is an n-simplex.

Definition 29.3. A **subcomplex** $K' \subset K$ of a simplicial complex $K = (V, S)$ is a simplicial complex $K' = (V', S')$ such that $V' \subset V$ and $S' \subset S$.

The **polyhedron** (*Polyeder*) of a simplicial complex $K = (V, S)$ is a topological space $|K|$ defined as follows. We denote by I^V the set of all functions $V \to I$, i.e. each element $t \in I^V$ is determined by a set of real numbers $t_v \in [0, 1]$ associated to the vertices $v \in V$, which we can think of as the *coordinates* of t. For each n-simplex $\sigma = \{v_0, \ldots, v_n\}$ in K, we define the set

$$|\sigma| := \left\{ t \in I^V \;\middle|\; \sum_{v \in \sigma} t_v = 1 \text{ and } t_v = 0 \text{ for all } v \notin \sigma \right\}.$$

This set is a copy of the standard n-simplex living in the finite-dimensional vector space $\mathbb{R}^\sigma \cong \mathbb{R}^{n+1}$, and we shall assign it the topology that it inherits naturally from this finite-dimensional vector space. As a set, the polyhedron $|K|$ is then defined by

$$|K| := \bigcup_{\sigma \in S} |\sigma| \subset I^V.$$

If K is finite, then $|K|$ lives inside the finite-dimensional vector space $\mathbb{R}^V$, and therefore has an obvious topology for which the topology we already defined on each of the subsets $|\sigma| \subset |K|$ matches the subspace topology. A bit more thought is required at this step if K is infinite. One possible choice would be to endow I^V with the product topology (via its obvious identification with $\prod_{v \in V} I$) and then take the subspace topology on $|K| \subset I^V$, but the product topology turns out not to be the most useful choice here. We will instead let the topology of $|K|$ be determined by that of the individual simplices:

Definition 29.4. Given a simplicial complex $K = (V, S)$, the topology of its polyhedron $|K| \subset I^V$ is defined such that a subset $\mathcal{U} \subset |K|$ is open if and only if $\mathcal{U} \cap |\sigma|$ is an open subset of $|\sigma|$ for every simplex $\sigma \in S$.

In other words, $|K|$ is equipped with the strongest[51] topology for which the inclusions $|\sigma| \hookrightarrow |K|$ are continuous for all σ. You should take a moment to convince yourself that this matches what

---

[51] For some unfathomable reason, the topology on $|K|$ has traditionally been referred to in the literature as the "weak" topology, and the same strange choice of nomenclature plagues the theory of CW-complexes, which we will discuss in a few weeks. It is a question of perspective: since $|K|$ has a lot of open sets, it is fairly difficult for sequences in $|K|$ to converge, or for maps into $|K|$ to be continuous, but on the flip side, it is relatively easy for functions defined on $|K|$ to be continuous (see Exercise 29.1).

was already said for the case where K is finite, and you should then prove the following proposition as an exercise:

Proposition 29.5. *For any simplicial complex $K = (V, S)$ and any space X, a map $f : |K| \to X$ is continuous if and only if $f|_{|\sigma|} : |\sigma| \to X$ is continuous for every simplex $\sigma \in S$.* □

Definition 29.6. A topological space X is a **polyhedron** (*Polyeder*) if it is homeomorphic to the polyhedron $|K|$ of some simplicial complex K. A choice of such a homeomorphism $X \cong |K|$ is called a **triangulation** (*Triangulierung*) or **simplicial decomposition** of the space X.

Remark 29.7. The definition of the term *triangulation* given above is perhaps stricter than some other sensible definitions of this term that one could imagine. What everyone can agree upon is that a triangulation of X should decompose X as a union of compact subsets, each of which is homeomorphic to a standard simplex, such that the intersection of any two of them is a common face of both; this includes the case where one of them is a face of the other, but also cases in which their interiors are disjoint. Definition 29.6 does decompose X in this way, but having a specific choice of homeomorphism $X \cong |K|$ is actually a lot more information, and it is debateable whether this amount of information is truly necessary for most of the important applications of triangulations. It will be useful for our purposes, however, when we want to write down precise relations between the simplicial and singular homologies of a triangulated space.

Definition 29.8. For each integer $n \geqslant 0$, the **n-skeleton** (*n-Skelett* or *n-Gerüst*) of a simplicial complex $K = (V, S)$ is the subcomplex $K^n = (V, S^n)$ of K whose set of simplices $S^n \subset S$ consists of all $\sigma \in S$ with $\dim \sigma \leqslant n$. Similarly, the **n-skeleton** of a polyhedron X with triangulation $X \cong |K|$ is the subspace $X^n \subset X$ formed by the polyhedron of the n-skeleton K^n of K.

This definition presents a polyhedron X as the union of a nested sequence of subspaces, its skeleta of various dimensions,

$$X^0 \subset X^1 \subset X^2 \subset \ldots \subset \bigcup_{n=0}^{\infty} X^n = X,$$

each of which is also a polyhedron. In particular, a polyhedron is n-dimensional (i.e. corresponds to an n-dimensional simplicial complex) if and only if it is equal to its n-skeleton. The 0-skeleton of any polyhedron is just the union of all its vertices—one can show that this is always a discrete set.

While $|K|$ was defined above as a subset of a vector space whose dimension may in general be quite large (or infinite), visualizing $|K|$ in concrete examples is often easier than one might expect.

Example 29.9. Suppose $V = \{v_0, v_1, v_2, v_3\}$ and S contains the subsets $A := \{v_0, v_1, v_2\}$ and $B := \{v_1, v_2, v_3\}$, plus all of their respective subsets. Then $|K|$ contains two copies of the triangle Δ^2, and they intersect each other along a single common edge connecting the vertices labeled v_1 and v_2. The complex is 2-dimensional, and its 1-skeleton is the union of all the edges of the triangles.

Example 29.10. If V has $n+1$ elements and S consists of all subsets of V except for V itself, then $|K|$ is homeomorphic to $\partial\Delta^n$, i.e. the union of all the boundary faces of Δ^n. In particular, this is homeomorphic to S^{n-1}.

Example 29.11. Suppose $V = \{v_0, \ldots, v_n\}$ for some $n \geqslant 2$ and S is defined to consist of all the one-element subsets $\{v_i\}$ plus the 1-simplices $\{v_i, v_{i+1}\}$ for $i = 0, \ldots, n-1$ and $\{v_n, v_0\}$. Then $|K|$ is a 1-dimensional polyhedron homeomorphic to S^1.

Example 29.12. Taking $V = \mathbb{Z}$ with S as the set of all 0-simplices $\{n\}$ plus 1-simplices of the form $\{n, n+1\}$ for $n \in \mathbb{Z}$ gives an infinite (but 1-dimensional) simplicial complex whose polyhedron is homeomorphic to $\mathbb{R}$.

Example 29.13. If $V = \mathbb{N}$ and S is the set of all finite subsets of $\mathbb{N}$, then K is an infinite-dimensional simplicial complex. Every simplex in this complex is a face of $\{1, \ldots, n\}$ for n sufficiently large, thus you can try to picture $|K|$ as the union of an infinite nested sequence of simplices $\Delta^0 \subset \Delta^1 \subset \Delta^2 \subset \ldots$, where each Δ^k is a boundary face of Δ^{k+1}.

Definition 29.14. Given two simplicial complexes $K_1 = (V_1, S_1)$ and $K_2 = (V_2, S_2)$, a **simplicial map** (*simpliziale Abbildung*) from K_1 to K_2 is a function $f : V_1 \to V_2$ such that $f(\sigma) \in S_2$ for every $\sigma \in S_1$.

Note that a simplicial map $K_1 \to K_2$ need not be injective on any given simplex, i.e. it can send an n-simplex of K_1 onto a k-simplex of K_2 for any $k \leqslant n$. There is a natural way to turn any simplicial map into a continuous map of the polyhedra $|K_1| \to |K_2|$. Indeed, denote by $\{e_v\}_{v \in V}$ the natural basis vectors of $\mathbb{R}^V$ so that every element $t \in \mathbb{R}^V$ can be written uniquely as a formal[52] sum $\sum_{v \in V} t_v e_v$ with coordinates $t_v \in \mathbb{R}$. Then since every element $t \in |K_1|$ is of the form $\sum_{v \in V_1} t_v e_v$ where only finitely many of the coordinates are nonzero and they all add up to 1, we can define

$$f : |K_1| \to |K_2| : \sum_{v \in V_1} t_v e_v \mapsto \sum_{v \in V_1} t_v e_{f(v)} \in I^{V_2}.$$

In other words, for each simplex $\sigma \in S_1$, f maps $|\sigma|$ onto $|f(\sigma)|$ via the restriction of the obvious linear map $\mathbb{R}^\sigma \to \mathbb{R}^{f(\sigma)}$ that sends basis vectors e_v to $e_{f(v)}$ for $v \in \sigma$. We have thus defined a functor

$$\mathsf{Simp} \to \mathsf{Top} : K \mapsto |K|,$$

where Simp is the category of simplicial complexes with morphisms defined to be simplicial maps. Notice that $f : |K_1| \to |K_2|$ always maps the n-skeleton of $|K_1|$ into the n-skeleton of $|K_2|$ for every $n \geqslant 0$.

Since we will often be concerned mainly with compact manifolds, the following result enables us to restrict attention to finite simplicial complexes:

Proposition 29.15. *A simplicial complex $K = (V, S)$ is finite if and only if its polyhedron $|K|$ is compact.*

This will follow from a more general theorem about CW-complexes that we shall prove in a few weeks, so for now, we'll settle for proving a special case, which happens to cover most of the interesting examples, and is quite easy:

Proof of Proposition 29.15 for finite-dimensional complexes. If K is finite, then $|K|$ is a closed and bounded subset of the finite-dimensional vector space $\mathbb{R}^V$, and is therefore compact.

Conversely, if K is infinite but $\dim K < \infty$, there exists an infinite sequence of distinct simplices $\sigma_1, \sigma_2, \ldots \in S$ with the property that each σ_i is not a face of any other simplex in K. Now for each $i \in \mathbb{N}$, pick a point $x_i \in |\sigma_i|$ along with an open neighborhood $\mathcal{U}_i \subset |\sigma_i|$ of x_i that is contained in the interior of $|\sigma_i|$. Since σ_i is not a face of any other simplex, we have $\mathcal{U}_i \cap |\sigma| = \varnothing$ for all simplices $\sigma \neq \sigma_i$, thus $\mathcal{U}_i$ defines an open subset of $|K|$ that contains x_i but none of the other points in the sequence $x_1, x_2, \ldots$. This proves that the infinite subset $\{x_1, x_2, \ldots\} \subset |K|$ is discrete, hence $|K|$ cannot be compact. □

---

[52] The word "formal" means in this context that we do not require the sum to converge in any sense, as it is a purely algebraic object. In practice, we are only going to consider points $t \in \mathbb{R}^V$ that have finitely many nonzero coordinates, thus the sums converge trivially.

29.2. The category of chain complexes. In absolute bordism theory, crucial roles are played by the words "closed" and "boundary": elements are represented by maps defined on *closed* manifolds rather than manifolds that are noncompact or have boundary, and the equivalence relation arises from the fact that certain manifolds are the boundaries of others. One trivial and yet important detail here is the fact that for a compact manifold M with boundary, its boundary ∂M is always a closed manifold: the compactness of ∂M is automatic since $\partial M \subset M$ is a closed subset, but being a boundary also means that ∂M cannot have any boundary points of its own.

In the usual constructions of homology theories—which do not require any knowledge of manifolds—there is an algebraic device that gives useful meaning to the words *closed* and *boundary*, and the fact that boundaries have no boundary of their own is then encoded by a simple algebraic equation, taking the form "$\partial^2 = 0$".

Definition 29.16. A **chain complex** (*Kettenkomplex*) of R-modules is a sequence of R-modules taking the form

$$\ldots \longrightarrow C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \longrightarrow \ldots$$

and satisfying the relation

$$\partial_n \circ \partial_{n+1} = 0 \tag{29.2}$$

for every $n \in \mathbb{Z}$.

Let's add some helpful terminology and notation to the definition above. The collection of R-modules C_n forming a chain complex can be packaged together as a single R-module

$$C_* := \bigoplus_{n \in \mathbb{Z}} C_n,$$

and writing $\partial : C_* \to C_*$ for the homomorphism determined uniquely by the maps $\partial_n : C_n \to C_{n-1}$ for all n, the defining relation (29.2) is then written succintly as

$$\partial^2 = 0.$$

The chain complex itself can then be denoted by (C_*, ∂), often abbreviated simply as C_*. We call ∂ the **boundary map** of **boundary operator** (*Randoperator*) of the complex. An element $c \in C_*$ is said to be **homogeneous** (*homogen*) if it belongs to the specific submodule $C_n \subset C_*$ for some $n \in \mathbb{Z}$, which is then called the **degree** (*Grad*) of c, sometimes written as

$$|c| := n \qquad \text{for } c \in C_n,$$

and the homogeneous elements of degree n are also called the n**-chains** (n-*Ketten*) of the complex. We say that $c \in C_*$ is **closed** (*geschlossen*) if it satisfies

$$\partial c = 0,$$

and the closed n-chains are called the n**-cycles** (n-*Zykel*) of the complex. Further, $c \in C_*$ is a **boundary** (*Rand*) if it satisfies

$$c = \partial a \qquad \text{for some} \qquad a \in C_*,$$

and the n-cycles that are also boundaries are called the n**-boundaries**. The relation $\partial^2 = 0$ is equivalent to the condition that all boundaries are also cycles, in other words, $\operatorname{im} \partial_{n+1}$ is always a submodule of $\ker \partial_n$.

Remark 29.17. For the boundary map $\partial : C_* \to C_*$ of a chain complex, one sometimes abuses notation and writes

$$\partial : C_* \to C_{*-1}$$

to emphasize the fact that ∂ sends n-chains to $(n-1)$-chains for each n. A fancier way to say this is that C_* is naturally a **$\mathbb{Z}$-graded** R-module, and the boundary map is a homomorphism of **degree** -1.

DEFINITION 29.18. The **homology** $H_*(C_*) = H_*(C_*, \partial)$ of the chain complex C_* is the collection of quotient modules

$$H_n(C_*) := \ker(\partial_n) \Big/ \operatorname{im}(\partial_{n+1}).$$

Their direct sum is denoted by

$$H_*(C_*) = \bigoplus_{n \in \mathbb{Z}} H_n(C_*).$$

Given a chain complex C_*, elements $[c] \in H_n(C_*)$ are called **homology classes** of degree n: their representatives $c \in C_n$ are n-cycles, and two such n-cycles c, c' represent the same homology class if and only if $c' - c$ is a boundary, in which case we say that they are **homologous**.

DEFINITION 29.19. Given two chain complexes (A_*, ∂^A) and (B_*, ∂^B), a **chain map** (*Kettenabbildung*) from (A_*, ∂^A) to (B_*, ∂^B) is a collection of homomorphisms $f_n : A_n \to B_n$ for $n \in \mathbb{Z}$ such that the following diagram commutes:

$$\begin{tikzcd}
\dots \arrow[r] & A_{n+1} \arrow[r, "\partial^A_{n+1}"] \arrow[d, "f_{n+1}"] & A_n \arrow[r, "\partial^A_n"] \arrow[d, "f_n"] & A_{n-1} \arrow[r, "\partial^A_{n-1}"] \arrow[d, "f_{n-1}"] & \dots \\
\dots \arrow[r] & B_{n+1} \arrow[r, "\partial^B_{n+1}"] & B_n \arrow[r, "\partial^B_n"] & B_{n-1} \arrow[r, "\partial^B_{n-1}"] & \dots
\end{tikzcd} \tag{29.3}$$

In other words, a chain map is a homomorphism $f : A_* \to B_*$ that maps n-chains to n-chains for each $n \in \mathbb{Z}$ and satisfies $\partial^B \circ f = f \circ \partial^A$.

It is easy to check that the composition of two chain maps is also a chain map, and so is the identity map on any chain complex, thus we can define a category

$$\mathsf{Ch}(R\text{-}\mathsf{Mod}) \qquad \text{often abbreviated as} \qquad \mathsf{Ch},$$

whose objects are chain complexes of R-modules, with chain maps as morphisms. The following easy observation then produces a functor

$$\mathsf{Ch}(R\text{-}\mathsf{Mod}) \xrightarrow{H_n} R\text{-}\mathsf{Mod}$$

for each $n \in \mathbb{Z}$, sending each chain complex to its homology in degree n and each chain map to the induced homomorphism between homologies.

PROPOSITION 29.20. *Any chain map $f : (A_*, \partial^A) \to (B_*, \partial^B)$ determines homomorphisms $f_* : H_n(A_*, \partial^A) \to H_n(B_*, \partial^B)$ for every $n \in \mathbb{Z}$ via the formula*

$$f_*[a] := [f(a)].$$

PROOF. There are two things to prove: first, that whenever $a \in A_n$ is a cycle, so is $f(a) \in B_n$. This is clear since $\partial^A a = 0$ implies $\partial^B(f(a)) = f(\partial^A a) = 0$ by the chain map condition. Second, we need to know that f maps boundaries to boundaries, so that it descends to a well-defined homomorphism $\ker \partial^A_n / \operatorname{im} \partial^A_{n+1} \to \ker \partial^B_n / \operatorname{im} \partial^B_{n+1}$. This is equally clear, since $a = \partial^A x$ implies $f(a) = f(\partial^A x) = \partial^B f(x)$. $\square$

29.3. Ordered simplicial homology. We now describe the first of two versions of the so-called **simplicial chain complex** (*simplizialer Kettenkomplex*) of a simplicial complex $K = (V, S)$, the homology of which will be the **simplicial homology** (*simpliziale Homologie*) of K. We will see later that with a bit of care, simplicial homology can be defined as a collection of functors on the subcategory of Top consisting of all polyhedra, without needing to specify how each polyhedron is triangulated. For now, however, the definition of the simplicial homology groups will depend explicitly on a simplicial complex, and thus gives us functors Simp → R-Mod.

The first version of the simplicial chain complex is algebraically simpler than the second, while the second will be easier to interpret geometrically. In practice, we will eventually be able to choose freely between them, because (for slightly nontrivial reasons) their homologies turn out to be naturally isomorphic.

Remark 29.21. For readers who have seen the definition of simplicial homology in the first semester of these notes (cf. Lecture 21): the complex defined in §29.4 below is cosmetically different from the one that was defined there, but is easily seen to be isomorphic to it (see Remark 29.24). The main difference is that our previous definition required fixing an arbitrary choice of orientation for each simplex, and the definition below avoids making any such choices.

Convention. For the rest of this lecture, and in fact for most of the rest of this course, you should assume that

$$G \in R\text{-}\mathsf{Mod}$$

is an arbitrary choice of R-module, which will typically play the role of the coefficient group in whichever version of homology is under discussion. We will include G in the notation for homology in situations where the choice of coefficient group matters, but omit it whenever this choice plays no important role.

Given a simplicial complex $K = (V, S)$, define the set

$$\mathcal{K}_n^o(K) := \left\{(v_0, \dots, v_n) \in V^{\times(n+1)} \;\middle|\; \text{there exists a } \sigma \in S \text{ with } v_i \in \sigma \text{ for all } i = 0, \dots, n\right\}$$

for each $n \geqslant 0$. The elements of $\mathcal{K}_n^o(K)$ are thus ordered $(n+1)$-tuples of vertices such that some simplex of the complex contains all of them. Note that in this definition, we are *not* assuming the $v_0, \dots, v_n$ are all distinct, though if they are, then it means $\{v_0, \dots, v_n\} \in S$ is an n-simplex of the complex K, and the ordered tuple $(v_0, \dots, v_n)$ is then called an **ordered n-simplex**. The **ordered simplicial chain complex** (with coefficients in G)

$$C_*^o(K) = C_*^o(K; G) = \bigoplus_{n \in \mathbb{Z}} C_n^o(K; G) = \bigoplus_{n \in \mathbb{Z}} C_n^o(K)$$

is defined with

$$C_n^o(K) = \bigoplus_{\sigma \in \mathcal{K}_n^o(K)} G$$

for each $n \geqslant 0$, so that n-chains can be written uniquely as finite sums $\sum_i a_i \sigma_i$ with coefficients $a_i \in G$ attached to canonical generators $\sigma_i \in \mathcal{K}_n^o(K)$. In particular, if the coefficient module G is taken to be the ring R itself, then $C_n^o(K)$ is the free R-module over the set $\mathcal{K}_n^o(K)$; in the case $R = \mathbb{Z}$, it is thus a free abelian group. Linearity and the formula

$$\partial(v_0, \dots, v_n) := \sum_{k=0}^{n} (-1)^k (v_0, \dots, v_{k-1}, v_{k+1}, \dots, v_n) \tag{29.4}$$

uniquely determine a boundary map $\partial : C_n^o(K) \to C_{n-1}^o(K)$ on this complex for each $n \geqslant 1$, and we define $C_n^o(K)$ to be trivial for each $n < 0$, so that $\partial : C_n^o(K) \to C_{n-1}^o(K)$ is necessarily trivial for each $n \leqslant 0$. It is a straightforward exercise in sign cancellations to verify that ∂ satisfies $\partial^2 = 0$. The resulting homology groups

$$H_*^o(K) := H_*^o(K;G) := H_*\big(C_*^o(K;G), \partial\big)$$

will be called the **ordered simplicial homology** of K with coefficients in G.

In order to view ordered simplicial homology as a functor, we associate to each simplicial map $f : K_1 \to K_2$ and each $n \geqslant 0$ the unique R-module homomorphism

$$f_* : C_n^o(K_1) \to C_n^o(K_2)$$

determined by linearity and the formula

$$f_*(v_0, \ldots, v_n) := (f(v_0), \ldots, f(v_n)).$$

It is straightforward to check that this defines a chain map $C_*^o(K_1) \to C_*^o(K_2)$, and thus gives us a functor

$$C_*^o : \mathsf{Simp} \to \mathsf{Ch}(R\text{-}\mathsf{Mod}).$$

Composing this with the algebraic homology functors $H_n : \mathsf{Ch}(R\text{-}\mathsf{Mod}) \to R\text{-}\mathsf{Mod}$ gives us functors

$$H_n^o : \mathsf{Simp} \to R\text{-}\mathsf{Mod};$$

in particular, simplicial maps $f : K_1 \to K_2$ induce R-module homomorphisms $f_* : H_n^o(K_1) \to H_n^o(K_2)$ for every n.

29.4. Oriented simplicial homology. The second version of the simplicial chain complex has a similar but smaller set of generators, because it excludes tuples $(v_0, \ldots, v_n)$ that contain repeats of the same vertex, and instead of keeping track of their orders, it keeps track of orientations. The following combinatorial result makes this possible; its proof is an exercise.

LEMMA 29.22. *For each $n \geqslant 1$, the boundary map $\partial : C_n^o(K;\mathbb{Z}) \to C_{n-1}^o(K;\mathbb{Z})$ defined via* (29.4) *preserves the subgroup of $C_*^o(K;\mathbb{Z})$ generated by all elements of the form*

$$\big(v_0, \ldots, v_n\big) \in C_n^o(K;\mathbb{Z}) \qquad \text{with } v_i = v_j \text{ for some } i \neq j \tag{29.5}$$

or of the form

$$\big(v_0, \ldots, v_n\big) - (-1)^{|\tau|}\big(v_{\tau(0)}, \ldots, v_{\tau(n)}\big) \in C_n^o(K;\mathbb{Z}) \tag{29.6}$$

for arbitrary $(v_0, \ldots, v_n) \in \mathcal{K}_n^o(K)$ and permutations $\tau \in S_{n+1}$, where $(-1)^{|\tau|} = \pm 1$ denotes the sign of the permutation. ☐

DEFINITION 29.23. An **orientation** (*Orientierung*) of an n-simplex $\sigma \in S$ for $n \geqslant 1$ in a complex $K = (V, S)$ is an equivalence class of orderings of the vertices of σ, where two orderings are considered equivalent if they differ by an even permutation. The case $n = 0$ is special: an orientation of a 0-simplex is simply a choice of sign $+1$ or -1, called the **positive** or **negative orientation** respectively.

A simplex endowed with an orientation is called an **oriented simplex** (*orientiertes Simplex*), and any oriented simplex with vertices $v_0, \ldots, v_n$ can be written with the notation

$$\pm[v_0, \ldots, v_n],$$

which is understood to mean the simplex $\{v_0, \ldots, v_n\}$ with orientation determined by the ordering $v_0, \ldots, v_n$ if the sign in front is positive, and the opposite of that orientation if the sign is negative. So for example, the symbols $[v_0, v_1]$ and $-[v_1, v_0]$ represent the same oriented 1-simplex, while that simplex with the opposite orientation can be written as either $-[v_0, v_1]$ or $[v_1, v_0]$. For an oriented 0-simplex $\pm[v_0]$, there is only one possible ordering, and the orientation is thus determined entirely by the initial sign. For a 2-simplex $\{v_0, v_1, v_2\}$, the fact that cyclic permutations of three elements are always even means

$$[v_0, v_1, v_2] = [v_1, v_2, v_0] = [v_2, v_0, v_1] = -[v_1, v_0, v_2] = -[v_0, v_2, v_1] = -[v_2, v_1, v_0].$$

In pictures of 2-dimensional polyhedra, one can usefully employ arrows on 1-simplices to specify orientations by ordering the two vertices, and circular arrows in 2-simplices to indicate the cyclic orderings that determine their orientations (see Figure 15).

Thanks to Lemma 29.22, the **oriented simplicial chain complex** of $K = (V, S)$ can be defined as a quotient

$$C_*^\Delta(K) := C_*^o(K)/D_*^o(K),$$

where we denote by $D_*^o(K) \subset C_*^o(K)$ the submodule generated by products of arbitrary coefficients $g \in G$ with elements of the form (29.5) or (29.6); this is sometimes called the group of **degenerate chains**. For each generator $(v_0, \ldots, v_n)$ of $C_n^o(K; \mathbb{Z})$, we shall denote the equivalence class that it represents in the quotient complex by

$$[v_0, \ldots, v_n] \in C_n^\Delta(K; \mathbb{Z}).$$

This means

$$[v_0, \ldots, v_n] = 0 \qquad \text{if } v_i = v_j \text{ for some } i \neq j,$$

whereas if the vertices $v_0, \ldots, v_n$ are all distinct, then $[v_0, \ldots, v_n]$ can be interpreted as an oriented n-simplex, and the equivalence relation in the quotient complex then reproduces our previous notational convention for oriented simplices, namely

$$[v_0, \ldots, v_i, \ldots, v_j, \ldots, v_n] = -[v_0, \ldots, v_j, \ldots, v_i, \ldots, v_n]$$

for each pair $i \neq j$ in $\{0, \ldots, n\}$. The boundary map $\partial : C_n^\Delta(K) \to C_{n-1}^\Delta(K)$ is thus determined by the formula

$$\partial[v_0, \ldots, v_n] = \sum_{k=0}^{n} (-1)^k [v_0, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n], \tag{29.7}$$

and Lemma 29.22 guarantees that this formula is independent of the order in which the vertices are written. We will denote the resulting **oriented simplicial homology** by

$$H_*^\Delta(K) := H_*\big(C_*^\Delta(K)\big).$$

One checks easily that the chain maps $f_* : C_*^o(K_1) \to C_*^o(K_2)$ induced by any simplicial map $f : K_1 \to K_2$ descend to the quotient as chain maps

$$f_* : C_*^\Delta(K_1) \to C_*^\Delta(K_2),$$

thus giving a functor

$$C_*^\Delta : \mathsf{Simp} \to \mathsf{Ch}(R\text{-}\mathsf{Mod}),$$

which composes with the algebraic homology functor to produce functors

$$H_n^\Delta : \mathsf{Simp} \to R\text{-}\mathsf{Mod}$$

for each $n \geqslant 0$.

Notice moreover that since the chain complex $C_*^\Delta(K)$ is defined as a quotient of $C_*^o(K)$, the quotient projection

$$C_*^o(K) \to C_*^\Delta(K) : (v_0, \ldots, v_n) \mapsto [v_0, \ldots, v_n]$$

is also a chain map, and thus induces a natural sequence of homomorphisms

$$H_n^o(K) \to H_n^\Delta(K), \qquad n \geqslant 0.$$

The word “natural” is meant here in its technical sense, as the map from ordered to oriented simplicial homology can be seen as a natural transformation between two functors $\mathsf{Simp} \to R\text{-}\mathsf{Mod}$.

We will later see in fact that on the homology level (though not on the level of chain complexes), these maps are always isomorphisms. This fact, however, requires a lengthier discussion, and we do not need it just yet.

Remark 29.24. While it is not so obvious from the definition above, $C_n^{\Delta}(K) = C_n^{\Delta}(K; G)$ for a simplicial complex K can be identified with a complex of the form

$$C_n^{\Delta}(K) \cong \bigoplus_{\sigma \in \mathcal{K}_n^{\Delta}(K)} G,$$

where the set $\mathcal{K}_n^{\Delta}(K)$ of generators consists of all n-simplices $\sigma = \{v_0, \ldots, v_n\}$ in the complex K. This perspective gives $C_*^{\Delta}(K)$ a similar formal structure to that of $C_*^o(K)$, so that for instance $C_n^{\Delta}(K; \mathbb{Z})$ is also a free abelian group, but with a smaller and more manageable set of generators than $C_n^o(K; \mathbb{Z})$. Indeed, each generator of $C_n^o(K)$ is an ordered tuple $(v_0, \ldots, v_n)$ of vertices in a simplex, but the generators of $C_n^{\Delta}(K)$ are instead actual n-simplices $\{v_0, \ldots, v_n\}$ of the complex K, meaning that the vertices $v_0, \ldots, v_n$ are required to be distinct, and the order in which they are written does not matter. Some choices are required, however, before $C_n^{\Delta}(K)$ can be presented in this way: if one makes an arbitrary choice of orientation for each simplex $\{v_0, \ldots, v_n\}$ of K and writes its vertices in an order consistent with the chosen orientation, then the resulting oriented n-simplex $[v_0, \ldots, v_n]$ can be used as a generator of $C_n^{\Delta}(K)$, and there is no need to consider other permutations of the vertices $v_0, \ldots, v_n$. Writing down $\partial : C_n^{\Delta}(K) \to C_{n-1}^{\Delta}(K)$ then requires taking some care with signs, to account for the fact that the arbitrarily chosen orientations of the $(n-1)$-simplices of K may or may not agree with the orientations of the boundary faces appearing in the usual formula for $\partial[v_0, \ldots, v_n]$. The result is essentially the definition of $H_*^{\Delta}(K)$ that we gave in Lecture 21 last semester, and it is also the description that typically seems most convenient for actual computations of simplicial homology (see e.g. Figure 15). The alternative formulation as a quotient complex shows why it does not actually depend on the choices of orientations.

Comparing the ordered and oriented simplicial chain complexes, the oriented complex has a more obvious geometric interpretation, because its generators are in bijective correspondence with actual simplices. By contrast, the ordered chain complex has a lot of redundant information, since each simplex gives rise to several generators corresponding to the different possible orderings of its vertices. But as we will see, the *ordered* complex is the one that admits a straightforward relationship with the singular homology of a polyhedron.

29.5. Exercises.

Exercise 29.1 (*). Prove Proposition 29.5: For any simplicial complex $K = (V, S)$ and any space X, a map $f : |K| \to X$ is continuous if and only if $f|_{|\sigma|} : |\sigma| \to X$ is continuous for every simplex $\sigma \in S$.

Exercise 29.2 (*). Prove Lemma 29.22, which establishes that the definition of ∂ on the oriented simplicial chain complex makes sense.
Hint: One does not really need to examine all possible tuples $(v_0, \ldots, v_n)$ and all of their permutations. It suffices to check cases where $v_k = v_{k+1}$ for some k, and permutations that interchange two neighboring elements.

Exercise 29.3. Figure 16 shows a simplicial complex $K = (V, S)$ whose associated polyhedron $|K|$ is homeomorphic to the Klein bottle. There are nine vertices labeled P_i, Q_i, R_i for $i = 1, 2, 3$, twenty-seven 1-simplices labeled by letters $a_i, b_i, c_i, d_i, e_i, f_i$ for $i = 1, 2, 3$ and g_i for $i = 1, \ldots, 9$, and eighteen 2-simplices labeled σ_i, τ_i for $i = 1, \ldots, 9$. The picture also shows a choice of orientation for each of the 2-simplices (circular arrows represent a cyclic ordering of the vertices) and 1-simplices (arrows point from the first vertex to the last). If we additionally endow each 0-simplex with the

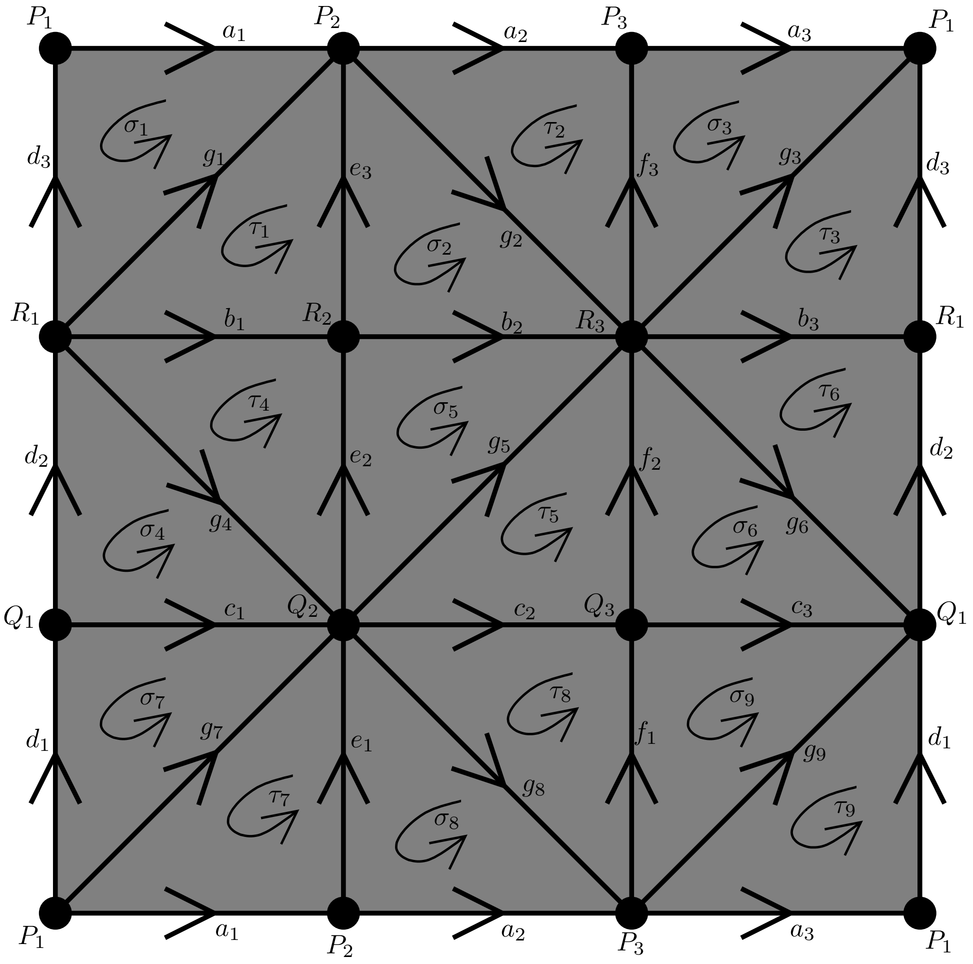


FIGURE 15. The picture shows a simplicial complex K with polyhedron $|K| \cong \mathbb{T}^2$, and choices of orientations on each simplex indicated via arrows (defining cyclic orderings of three vertices in the case of each 2-simplex). With these orientations fixed, plugging in the definition of $\partial : C_n^\Delta(K;\mathbb{Z}) \to C_{n-1}^\Delta(K;\mathbb{Z})$ gives e.g. $\partial\sigma_1 = g_1 - a_1 - d_3$, $\partial\tau_1 = b_1 + e_3 - g_1$, $\partial a_1 = P_2 - P_1$, $\partial a_2 = P_3 - P_2$, $\partial a_3 = P_1 - P_3$, and so forth. The complete computation of $H_*^\Delta(K;\mathbb{Z})$ was carried out near the end of Lecture 21 last semester, with $H_2^\Delta(K;\mathbb{Z}) \cong \mathbb{Z}$ generated by the sum of the eighteen 2-simplices in the complex, $H_1^\Delta(K;\mathbb{Z}) \cong \mathbb{Z}^2 \cong \pi_1(\mathbb{T}^2) \cong H_1(\mathbb{T}^2;\mathbb{Z})$, and $H_0^\Delta(K;\mathbb{Z}) \cong \mathbb{Z} \cong H_0(\mathbb{T}^2;\mathbb{Z})$.

positive orientation, every letter in the picture can be regarded as representing an oriented simplex, and thus a generator of the oriented simplicial chain complex $C_*^\Delta(K;\mathbb{Z})$.

(a) Write down the 1-chains $\partial\sigma_i, \partial\tau_i \in C_1^\Delta(K;\mathbb{Z})$ explicitly for each $i = 1, \dots, 9$.
(b) Prove that $H_2^\Delta(K;\mathbb{Z}_2) \cong \mathbb{Z}_2$, and write down a specific cycle in $C_2^\Delta(K;\mathbb{Z}_2)$ that generates it.
(c) Prove that $H_2^\Delta(K;\mathbb{Z}) = 0$.

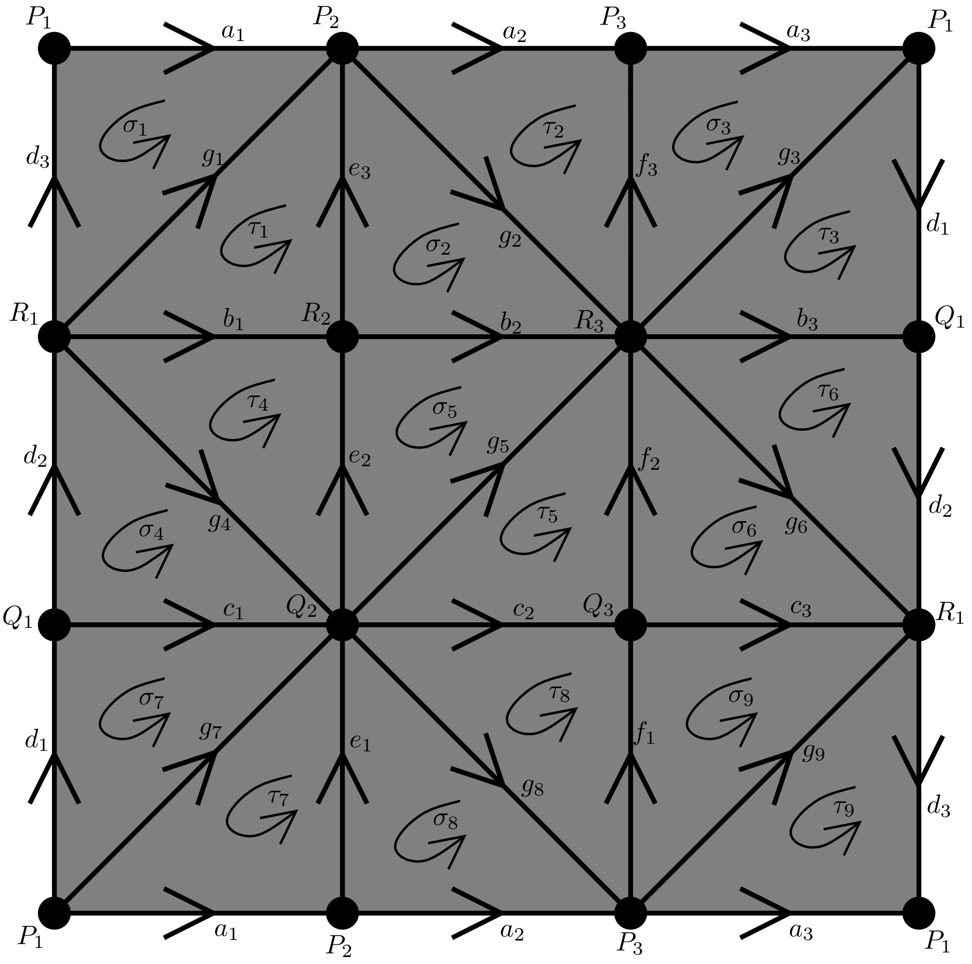


FIGURE 16. The Klein bottle as a polyhedron.

(d) Show that the 1-cycle $d_1 + d_2 + d_3$ represents a nontrivial homology class $[d_1 + d_2 + d_3]$ in both $H_1^\Delta(K;\mathbb{Z})$ and $H_1^\Delta(K;\mathbb{Z}_2)$, but satisfies $2[d_1 + d_2 + d_3] = 0 \in H_1^\Delta(K;\mathbb{Z})$ and $[d_1 + d_2 + d_3] = 0 \in H_1^\Delta(K;\mathbb{Q})$.

EXERCISE 29.4 (*). The following computations may give you a hint as to why $h_0(\{*\})$ is called the *coefficient group* of an axiomatic homology theory h_*. In the simplicial context, let $\{*\}$ denote a simplicial complex that has exactly one vertex.

(a) Prove that $H_0^\Delta(\{*\};G) \cong G$ and $H_n^\Delta(\{*\};G) = 0$ for all $n \neq 0$.
Hint: This is nearly trivial.

(b) Prove that $H_0^o(\{*\};G) \cong G$ and $H_n^o(\{*\};G) = 0$ for all $n \neq 0$.
Remark: This is slightly less trivial than part (a), but not difficult.

30. Triangulated manifolds and subdivision

I claimed in the previous lecture that simplicial homology can be viewed as a combinatorial variant of bordism theory. To see what I mean by this, we need to talk about manifolds with triangulations.

30.1. Triangulated manifolds. In this lecture, we will not need any knowledge of smoothness, so the word "manifold" means *topological* manifold, i.e. a second countable Hausdorff space that is locally homeomorphic to a finite-dimensional vector space or half-space. It should be assumed that all manifolds M may have nonempty boundary ∂M unless stated otherwise.

Definition 30.1. An n-dimensional **triangulated manifold** is a topological n-manifold M equipped with a triangulation $M \cong |K|$ that identifies ∂M with the polyhedron of a subcomplex $K' \subset K$.

The following is a consequence of the local Euclidean structure of manifolds:

Proposition 30.2. *If $M \cong |K|$ is a triangulated n-dimensional manifold, then the associated simplicial complex K is n-dimensional, and every $(n-1)$-simplex σ in K is a boundary face of either one or two n-simplices, where the former is the case if and only if σ belongs to the subcomplex triangulating ∂M.* □

In general, it is a subtle question whether a given manifold admits a triangulation. It is known to be true for all *smooth* manifolds, and also for topological manifolds of dimension at most three (see [Moi77]), but not in general for dimensions four and above (see [Man14]). We will not concern ourselves with such questions here, as for our purposes, it is already helpful to consider explicit examples of manifolds with triangulations, such as the picture of $\mathbb{T}^2$ in Figure 15. Our immediate motivation for doing so is to give explicit constructions of some important homology classes. The idea is to turn a triangulation $M \cong |K|$ of a compact n-manifold into an n-chain in the simplicial chain complex of K.

It is easiest to explain how this works in $H^{\Delta}(K;\mathbb{Z}_2)$. Using $\mathbb{Z}_2$ as a coefficient group has the advantage that for any n-simplex $\sigma = \{v_0, \dots, v_n\}$ of K, we have

$$[v_0, \dots, v_n] = -[v_0, \dots, v_n] \in C_n^{\Delta}(K;\mathbb{Z}_2),$$

so that all choices of ordering for the vertices $v_0, \dots, v_n$ produce the same element, and there is thus no need to worry about orientations. Given a compact triangulated n-manifold $M \cong |K|$, we can define an oriented simplicial n-chain by

$$c_M := \sum_{\sigma} \mathbf{v}_\sigma \in C_n^{\Delta}(K;\mathbb{Z}_2), \tag{30.1}$$

where the sum ranges over the set of all n-simplices σ of K, and $\mathbf{v}_\sigma = [v_0, \dots, v_n]$ denotes the vertices of $\sigma = \{v_0, \dots, v_n\}$, arranged in an arbitrary order. Note that this definition would not make sense if M were not compact, but according to Proposition 29.15, compactness implies that K is a *finite* simplicial complex, so that the sum in the definition of c_M is finite. If $\partial M \neq \varnothing$, then the subcomplex $K' \subset K$ triangulating ∂M similarly defines a simplicial $(n-1)$-chain

$$c_{\partial M} \in C_{n-1}^{\Delta}(K';\mathbb{Z}_2) \subset C_{n-1}^{\Delta}(K;\mathbb{Z}_2),$$

where we are regarding $C_{n-1}^{\Delta}(K';\mathbb{Z}_2)$ as a submodule of $C_{n-1}^{\Delta}(K;\mathbb{Z}_2)$, which makes sense because the canonical generators of $C_{n-1}^{\Delta}(K';\mathbb{Z}_2)$ (i.e. the $(n-1)$-simplices of K') are also $(n-1)$-simplices of K and thus generators of $C_{n-1}^{\Delta}(K;\mathbb{Z}_2)$. If $\partial M = \varnothing$, then the recipe above defines the trivial $(n-1)$-chain, and we can therefore sensibly write

$$c_{\partial M} = 0 \in C_{n-1}^{\Delta}(K;\mathbb{Z}_2) \qquad \text{if } \partial M = \varnothing.$$

PROPOSITION 30.3. *The chains* $c_M \in C_n^\Delta(K;\mathbb{Z}_2)$ *and* $c_{\partial M} \in C_{n-1}^\Delta(K';\mathbb{Z}_2) \subset C_{n-1}^\Delta(K;\mathbb{Z}_2)$ *in the situation above satisfy*

$$\partial c_M = c_{\partial M}.$$

PROOF. By Proposition 30.2, applying ∂ to the right hand side of (30.1) produces exactly two copies of each $(n-1)$-simplex of K that is not in K', so with $\mathbb{Z}_2$ coefficients, they cancel each other. What remains is a single term for each $(n-1)$-simplex in the triangulation of ∂M, which produces $c_{\partial M}$. □

The proposition implies in particular that whenever M is a *closed* triangulated n-manifold, the n-chain c_M is a cycle, and thus represents a homology class

$$[M] := [c_M] \in H_n^\Delta(K;\mathbb{Z}_2).$$

We call this the (simplicial) **fundamental class** of M, and refer to c_M as a **fundamental cycle**. In the case $\partial M \neq \varnothing$, we will see when we discuss *relative* simplicial homology that c_M still represents a relative homology class for the triangulated pair of spaces $(M, \partial M)$, thus the terms *fundamental cycle* and *fundamental class* remain appropriate.

Fundamental cycles and classes can also be defined in ordered simplicial homology, but this requires some choices.

DEFINITION 30.4. An **admissible ordering** on a simplicial complex $K = (V, S)$ assigns to each simplex $\sigma \in S$ a total order on its set of vertices such that the inclusion $\tau \hookrightarrow \sigma$ of each of its faces $\tau \subset \sigma$ is an order-preserving map.

It is easy to see that every simplicial complex admits an admissible ordering, e.g. one can simply choose a total order on the entire set of vertices V, and define the total orders on every simplex $\sigma \subset V$ so that the inclusion $\sigma \hookrightarrow V$ is order preserving. Since we are only talking about compact manifolds in this lecture, our simplicial complexes are always finite, so you don't even need to appeal to any abstract set-theoretic machinery (e.g. the axiom of choice) before choosing a total order on V. There are also situations where establishing a rule to determine total orders on every simplex $\sigma \in S$ is more convenient than choosing a total order on V itself.

Suppose again that $M \cong |K|$ is a compact triangulated n-manifold, and let $K' \subset K$ denote the subcomplex whose polyhedron is identified with ∂M. Working with $\mathbb{Z}_2$ coefficients, any choice of admissible ordering for K determines an ordered simplicial n-chain of the form

$$c_M := \sum_\sigma \mathbf{v}_\sigma \in C_n^o(K;\mathbb{Z}_2),$$

in which the sum ranges again over the set of all n-simplices σ of K, and $\mathbf{v}_\sigma = (v_0, \dots, v_n)$ denotes the vertices of $\sigma = \{v_0, \dots, v_n\}$ arranged in increasing order. If $\partial M \neq \varnothing$, the admissible ordering on K restricts to an admissible ordering on K', and thus similarly determines an ordered simplicial $(n-1)$-chain

$$c_{\partial M} \in C_{n-1}^o(K';\mathbb{Z}_2) \subset C_{n-1}^o(K;\mathbb{Z}_2),$$

and we take $c_{\partial M}$ to be $0 \in C_{n-1}^o(K;\mathbb{Z}_2)$ if $\partial M = \varnothing$. It is easy to verify that the analogue of Proposition 30.3 also holds in this situation, and we thus have

$$\partial c_M = c_{\partial M} \in C_{n-1}^o(K;\mathbb{Z}_2).$$

It is clear from the construction that the natural chain map $C_*^o(K;\mathbb{Z}_2) \to C_*^\Delta(K;\mathbb{Z}_2)$ sends each ordered fundamental cycle to the oriented fundamental cycle, so when M is closed, it therefore sends an ordered simplicial fundamental class $[M] \in H_n^o(K;\mathbb{Z}_2)$ to the oriented simplicial fundamental class $[M] \in H_n^\Delta(K;\mathbb{Z}_2)$. Once we've proved that the natural map $H_n^o(K;\mathbb{Z}_2) \to H_n^\Delta(K;\mathbb{Z}_2)$ is an isomorphism, we will be able to deduce from this that the class $[M] \in H_n^o(K;\mathbb{Z}_2)$ is independent of choices, even though the fundamental cycle $c_M \in C_n^o(K;\mathbb{Z}_2)$ that represents it does depend on the choice of admissible ordering for the complex.

The following result is a worthwhile exercise in the computation of simplicial homology.

THEOREM 30.5. *For any closed and connected triangulated n-manifold $M \cong |K|$, $H_n^\Delta(K;\mathbb{Z}_2)$ is isomorphic to $\mathbb{Z}_2$, and its unique nontrivial element is the fundamental class $[M]$.* □

30.2. Oriented triangulations. In order to extend the construction of fundamental cycles from $\mathbb{Z}_2$ to integer coefficients, we need triangulations with a bit of extra structure.

DEFINITION 30.6. Suppose $n \geqslant 1$ and $\pm[v_0,\dots,v_n]$ is an oriented n-simplex in a simplicial complex. The induced **boundary orientation** on the boundary face $\{v_1,\dots,v_n\}$ is then given by the oriented $(n-1)$-simplex $\pm[v_1,\dots,v_n]$.

Note that the oriented simplex $\pm[v_0,\dots,v_n]$ can typically be written in multiple distinct ways with the vertex v_0 appearing first and the other vertices permuted, but the same permutation then applies to the oriented boundary face $\pm[v_1,\dots,v_n]$ and causes the same sign change, so that Definition 30.6 does not depend on any choices. Moreover, the definition determines an orientation on *every* boundary face of $\sigma = \{v_0,\dots,v_n\}$, because for any $k = 0,\dots,n$, one can always apply a permutation to rewrite $\pm[v_0,\dots,v_n]$ with v_k in front; in particular, $[v_0,\dots,v_n] = (-1)^k[v_k,v_0,\dots,v_{k-1},v_{k+1},\dots,v_n]$, so that endowing the face $\{v_0,\dots,v_{k-1},v_{k+1},\dots,v_n\}$ with the boundary orientation determined by $[v_0,\dots,v_n]$ produces the oriented simplex

$$(-1)^k[v_0,\dots,v_{k-1},v_{k+1},\dots,v_n].$$

The formula (29.7) for $\partial[v_0,\dots,v_n]$ in the oriented simplicial chain complex can thus be interpreted as the sum of the $n+1$ boundary faces of $[v_0,\dots,v_n]$ endowed with their boundary orientations.

REMARK 30.7. In addition to being consistent with our usual formulas for boundary operators on chain complexes, there is some geometric motivation behind Definition 30.6. In differential geometry, an oriented n-manifold M induces a natural boundary orientation on ∂M, and if M has a triangulation, the orientation of M also induces orientations of the n-simplices in its triangulation. One can check that if the polyhedron $|\sigma|$ of an oriented n-simplex σ in a complex K is viewed as an oriented n-manifold, then the geometric notion of boundary orientation on $\partial|\sigma| \cong S^{n-1}$ matches the induced orientations (according to Definition 30.6) of the boundary faces of σ, which form a triangulation of $\partial|\sigma|$.

DEFINITION 30.8. For an n-dimensional manifold M, an **oriented triangulation** (*orientierte Triangulierung*) of M is a triangulation in which every n-simplex is endowed with an orientation such that for every $(n-1)$-simplex σ not contained in ∂M, the two boundary orientations it inherits as a boundary face of two distinct oriented n-simplices (cf. Prop. 30.2) are opposite.

I recommend now taking another look at Figure 15 to verify that the orientations of 2-simplices depicted in this picture define an oriented triangulation of $\mathbb{T}^2$. Then, contrast it with Figure 16, which shows a triangulation of the Klein bottle in which orientations of the 2-simplices have been chosen but they fail to satisfy the conditions of Definition 30.8. (The trouble is with the 1-simplices labeled d_1, d_2, d_3.) The problem with the Klein bottle is of course that it is a *non-orientable* manifold, and it turns out that only orientable manifolds can admit oriented triangulations—we sketched a proof of this for surfaces last semester in Lecture 20, and we will be able to prove it for all manifolds later in this course using homology.

EXAMPLE 30.9. The triangulation of S^{n-1} described in Example 29.10 can be oriented by choosing an ordering of the vertex set V, regarding this as an oriented n-simplex σ and then endowing each of its boundary faces with the boundary orientation. The cancelation condition on $(n-2)$-simplices in this case is roughly equivalent to the fact that $\partial^2 = 0$ in the singular and simplicial chain complexes; see Proposition 30.10 below.

We now consider whether a version of the fundamental cycle $c_M \in C_n^\Delta(K;\mathbb{Z}_2)$ for a compact triangulated n-manifold $M \cong |K|$ with boundary $\partial M \cong |K'|$ can also be defined with integer coefficients. Indeed, suppose that an orientation has been chosen for each of the n-simplices σ of K, and consider an n-chain of the form

$$c_M := \sum_\sigma \mathbf{v}_\sigma \in C_n^\Delta(K;\mathbb{Z}), \tag{30.2}$$

where as usual the sum ranges over the set of all n-simplices $\sigma = \{v_0, \ldots, v_n\}$ in K, and $\mathbf{v}_\sigma = [v_0, \ldots, v_n]$ is defined by ordering the vertices in accordance with the chosen orientation. Since each $(n-1)$-simplex of the subcomplex $K' \subset K$ triangulating ∂M is a boundary face of a unique n-simplex, the chosen orientations of the n-simplices determine boundary orientations of the $(n-1)$-simplices of K', which we can use to define an $(n-1)$-chain

$$c_{\partial M} \in C_{n-1}^\Delta(K';\mathbb{Z}) \subset C_{n-1}^\Delta(K;\mathbb{Z}).$$

The formula

$$\partial c_M = c_{\partial M}$$

is then satisfied if and only if the chosen orientations of the n-simplices satisfy the condition in Definition 30.8: indeed, this condition means that all contributions to ∂c_M from $(n-1)$-simplices not in ∂M appear in cancelling pairs, while each $(n-1)$-simplex in ∂M appears exactly once with the correct sign. In the case $\partial M = \emptyset$, $c_M \in C_n^\Delta(K;\mathbb{Z})$ is then a cycle and thus represents an integral fundamental class

$$[M] := [c_M] \in H_n^\Delta(K;\mathbb{Z}).$$

We summarize:

PROPOSITION 30.10. *For any compact triangulated n-manifold $M \cong |K|$ with an oriented triangulation, the induced triangulation of the boundary $\partial M \cong |K'|$ admits a unique orientation for which each $(n-1)$-simplex of K' is oriented as the boundary of an oriented n-simplex of K. The resulting fundamental cycles in $C_*^\Delta(K;\mathbb{Z})$ as constructed above then satisfy the relation $\partial c_M = c_{\partial M}$.*

PROOF. The discussion preceding the statement showed that if the triangulation of M is oriented and the orientations of its n-simplices are used in defining $c_M \in C_n^\Delta(K;\mathbb{Z})$ and (via boundary orientations) $c_{\partial M} \in C_{n-1}^\Delta(K';\mathbb{Z}) \subset C_{n-1}^\Delta(K;\mathbb{Z})$, then $\partial c_M = c_{\partial M}$. One detail not yet addressed is that the boundary orientations on the $(n-1)$-simplices of K' really do define an oriented triangulation of ∂M: this follows from the relation

$$\partial c_{\partial M} = \partial(\partial c_M) = 0,$$

which means that the two contributions to $\partial c_{\partial M}$ from each $(n-2)$-simplex in ∂M cancel each other. ☐

Here is another worthwhile computational exercise:

THEOREM 30.11. *For any closed and connected n-manifold $M \cong |K|$ with an oriented triangulation, $H_n^\Delta(K;\mathbb{Z})$ is isomorphic to $\mathbb{Z}$, and its fundamental class $[M]$ is a generator.* ☐

For an analogue in ordered simplicial homology with integer coefficients, we can again choose an admissible ordering for K, but we need to be aware that the resulting ordering of the vertices $(v_0, \ldots, v_n)$ of each simplex σ might not be consistent with the chosen orientation of σ. We can account for this by including appropriate signs in the formula: we define

$$c_M := \sum_\sigma \epsilon_\sigma \mathbf{v}_\sigma \in C_n^o(K;\mathbb{Z}),$$

where for each n-simplex $\sigma = \{v_0, \dots, v_n\}$ of our oriented triangulation, with vertices arranged in increasing order, we set $\epsilon_\sigma = \pm 1$ so that $\epsilon_\sigma[v_0, \dots, v_n]$ defines the chosen orientation. Defining

$$c_{\partial M} \in C^o_{n-1}(K'; \mathbb{Z}) \subset C^o_{n-1}(K; \mathbb{Z})$$

in the same manner via the admissible ordering and boundary orientation, the same arguments as before prove $\partial c_M = c_{\partial M}$.

30.3. Triangulated bordism. With triangulated manifolds in hand, the similarity between homology and bordism theory can be made more explicit. Suppose

$$X \cong |K|$$

is a space (but not necessarily a manifold) triangulated by a simplicial complex K, and suppose

$$M \cong |L|$$

is a closed triangulated n-manifold with an oriented triangulation. Any simplicial map $\varphi : L \to K$ induces a continuous map $\varphi : M \to X$, so that the pair (M, φ) represents an element of the oriented bordism group $\Omega^{\mathrm{SO}}_n(X)$. A corresponding simplicial homology class can be defined by

$$\varphi_*[M] \in H^\Delta_n(K; \mathbb{Z}),$$

using the integral fundamental class $[M] = [c_M] \in H^\Delta_n(L; \mathbb{Z})$ defined via the oriented triangulation of M. Further, suppose there is an oriented bordism (W, Φ) between (M, φ) and another such pair (N, ψ), equipped with the additional data of an oriented triangulation: more precisely, $W \cong |L'|$ is a compact $(n+1)$-manifold with an oriented triangulation, $\Phi : L' \to K$ is a simplicial map inducing the continuous map $\Phi : W \to X$, and there is a homeomorphism $\partial W \cong M \amalg (-N)$ that identifies the subcomplex triangulating M with K and $\Phi|_M$ with φ. The minus sign in front of N means that we consider $N \subset \partial W$ to be equipped with an oriented triangulation whose n-simplices carry the *opposite* of the boundary orientations they inherit from the oriented $(n+1)$-simplices triangulating W. With this understood, we can write $\psi := \Phi|_N : N \to X$ and obtain a simplicial homology class

$$\psi_*[N] \in H^\Delta_n(K; \mathbb{Z})$$

in the same manner as $\varphi_*[M]$, but the triangulated bordism (W, Φ) tells us more: the orientation reversal on N gives the relation

$$\partial c_W = c_{\partial W} = c_M - c_N \in C^\Delta_n(L'; \mathbb{Z}),$$

and plugging this into the chain map $\Phi_* : C^\Delta_*(L'; \mathbb{Z}) \to C^\Delta_*(K; \mathbb{Z})$, we have

$$\partial(\Phi_* c_W) = \Phi_*(c_M - c_N) = \varphi_* c_M - \psi_* c_N \in C^\Delta_n(K; \mathbb{Z}),$$

implying

$$\varphi_*[M] = \psi_*[N] \in H^\Delta_n(K, \mathbb{Z}).$$

This matches what happens in bordism theory: two simplicial homology classes represented by closed triangulated manifolds with simplicial maps are the same whenever there is a *triangulated* bordism between them. We will see when we study fundamental classes in singular homology that the entire discussion makes sense in that context as well, but without any need for triangulations.

Remark 30.12. Orientations were needed for all the triangulations in the discussion above because we were working with integer coefficients. If we did not have orientations, the entire discussion would still make sense after uniformly replacing the coefficient group $\mathbb{Z}$ by $\mathbb{Z}_2$, and $H^\Delta_n(X; \mathbb{Z}_2)$ thus becomes the combinatorial variant of the *unoriented* bordism group $\Omega^{\mathrm{O}}_n(X)$.

30.4. Barycentric subdivision. I would now like to describe a specific triangulation of the standard n-simplex Δ^n. One can reasonably ask why it is worth bothering to triangulate a simplex: after all, Δ^n is already a polyhedron in a trivial way. But the point of the following construction is to decompose Δ^n into n-simplices that are strictly *smaller*, and iterating the process will then produce triangulations whose individual n-simplices are as small as we like. This will come in handy when we need to prove the formal properties of singular homology, and it also has some important theoretical consequences for simplicial homology, including one ingredient in the proof that $H_*^o(K)$ and $H_*^\Delta(K)$ really are invariants of the *polyhedron* $|K|$, and not just of the particular simplicial complex K that is used for triangulating it.

For each $n \geqslant 0$, the point

$$b_n := \left(\frac{1}{n+1}, \dots, \frac{1}{n+1}\right) \in \Delta^n \subset \mathbb{R}^{n+1}$$

is called the **barycenter** of the standard n-simplex; you should imagine it as the center of mass of Δ^n. The following inductive procedure uniquely determines a decomposition of Δ^n for each $n \geqslant 0$ into smaller pieces $\delta^n \subset \Delta^n$ that are homeomorphic to Δ^n:

(1) For $n = 0$, the one-point space Δ^0 cannot be decomposed any further, so its triangulation consists only of a single 0-simplex.

(2) If the triangulation of Δ^{n-1} has already been defined, then using the canonical identification of the boundary face $\partial_{(k)}\Delta^n$ for each $k = 0, \dots, n$ with Δ^{n-1}, each $(n-1)$-simplex $\delta^{n-1} \subset \partial_{(k)}\Delta^n$ in its triangulation determines an n-simplex $\delta^n \subset \Delta^n$ as the convex hull of δ^{n-1} and the barycenter b_n.

For a more precise description of barycentric subdivision, we should specify an abstract simplicial complex K along with a homeomorphism $|K| \cong \Delta^n$ defining the triangulation of Δ^n. It is most natural to define K so that its vertices are points in Δ^n, and since Δ^n is a subset of the vector space $\mathbb{R}^{n+1}$, the following condition becomes relevant:

Definition 30.13. Given a vector space V of dimension at least n, a set of n points in V is said to be in **general position** if they are not contained in any $(n-2)$-dimensional plane.

For example, three points in a vector space of dimension at least 2 are in general position if they are not colinear. In general, for a given set of points $v_0, \dots, v_n \in V$ with $\dim V \geqslant n+1$, the unique linear map $\mathbb{R}^{n+1} \to V$ sending the standard basis of $\mathbb{R}^{n+1}$ to the vectors $v_0, \dots, v_n$ restricts to $\Delta^n \subset \mathbb{R}^{n+1}$ as an *embedding* $\Delta^n \hookrightarrow V$ if and only if the points $v_0, \dots, v_n$ are in general position.

The abstract simplicial complex K arising from the barycentric subdivision of Δ^n can now be described as follows. One first triangulates the standard 0-simplex with the simplicial complex whose only vertex is the one point in $\Delta^0 \subset \mathbb{R}$. Then inductively, having defined triangulations of the boundary faces $\partial_{(k)}\Delta^n \cong \Delta^{n-1}$ via complexes whose vertices are all identified with points in $\partial_{(k)}\Delta^n$, each n-simplex of K is defined to have vertices $b_n, v_1, \dots, v_n$, where $v_1, \dots, v_n \in \partial_{(k)}\Delta^n$ are the vertices of an $(n-1)$-simplex in the complex triangulating $\partial_{(k)}\Delta^n \cong \Delta^{n-1}$ for some $k = 0, \dots, n$. The homeomorphism identifying $|K|$ with Δ^n sends each simplex $|\sigma| \subset |K|$ into Δ^n via the restriction of the unique linear map that sends each vertex to itself. That this actually defines a homeomorphism $|K| \cong \Delta^n$ follows from the following proposition, whose proof is Exercise 30.2:

Proposition 30.14. *For the abstract simplicial complex K described above, whose vertices are points in Δ^n:*

(a) The vertices of each n-simplex are in general position.

(b) Every point $p \in \Delta^n$ lies in the convex hull of the points $v_0, \dots, v_k \in \Delta^n$ for some simplex $\{v_0, \dots, v_k\}$ of K.

□

It is worth pausing a moment now to draw pictures of the barycentric subdivisions of Δ^n for $n = 1, 2$: the subdivision of Δ^1 will have only two 1-simplices, and since Δ^2 has three boundary faces, its subdivision has six 2-simplices. In general, the number of n-simplices in the subdivision of Δ^n will be $(n+1)!$. You'll find a picture of the subdivision of Δ^3 in [**Hat02**], among other places; if you ever find a convincing picture of the case $n = 4$, let me know.

The triangulation defined above can be turned into an integral fundamental cycle

$$c_{\Delta^n} \in C_n^{\Delta}(K;\mathbb{Z}) \qquad \text{or} \qquad c_{\Delta^n} \in C_n^{o}(K;\mathbb{Z})$$

after making some additional choices, namely of orientations and/or admissible orderings. There is surely more than one possible recipe for this, but here is one that works. Inductively, assume an admissible ordering and an orientation have already been chosen for the barycentric subdivision of Δ^{n-1}; for the case $n = 0$, there is no choice of ordering to be made, and we can fix the positive orientation on the unique 0-simplex. Now if $v_1, \ldots, v_n$ are the vertices of an $(n-1)$-simplex on one of the boundary faces $\partial_{(k)}\Delta^n = \Delta^{n-1}$ arranged in increasing order, and $\pm[v_1, \ldots, v_n]$ is its chosen orientation, define the ordering and orientation of the n-simplex $\{b_n, v_1, \ldots, v_n\}$ in Δ^n to be given by

$$(b_n, v_1, \ldots, v_n) \qquad \text{and} \qquad \pm(-1)^k[b_n, v_1, \ldots, v_n]$$

respectively. Following our usual prescriptions to define fundamental cycles as simplicial n-chains in $C_n^o(K;\mathbb{Z})$ or $C_n^{\Delta}(K;\mathbb{Z})$, the barycentric subdivisions of Δ^n and its boundary faces are then related via the formula

$$\partial c_{\Delta^n} = \sum_{k=0}^{n} (-1)^k c_{\partial_{(k)}\Delta^n} \qquad \text{in} \qquad C_{n-1}^{o}(K;\mathbb{Z}) \text{ or } C_{n-1}^{\Delta}(K;\mathbb{Z}),$$

where as usual, $c_{\partial_{(k)}\Delta^n}$ is defined by identifying $\partial_{(k)}\Delta^n$ with Δ^{n-1} and is then regarded as an element of $C_{n-1}^{\Delta}(K;\mathbb{Z})$ since the vertices of the subdivision of $\partial_{(k)}\Delta^n$ are also vertices of the subdivision of Δ^n.

Since we can now subdivide a standard simplex into smaller simplices of the same dimension, we can also subdivide *any* polyhedron. Indeed, assuming $K = (V, S)$ is an arbitrary simplicial complex, for each simplex $\sigma \in S$ of K, the corresponding subset $|\sigma| \subset |K|$ has a well-defined barycenter $b_\sigma \in |\sigma|$. We can then construct a new simplicial complex $K' = (V', S')$ whose vertices are points in the polyhedron $|K|$, including all the vertices of K plus all the barycenters of its simplices, and such that the simplices of K' correspond to simplices in the barycentric subdivisions of the individual simplices of K. The unique linear map $\mathbb{R}^{V'} \to \mathbb{R}^V$ that sends the basis vector $e_v \in \mathbb{R}^{V'}$ corresponding to each vertex $v \in V'$ to the location of that vertex in $|K| \subset \mathbb{R}^V$ now restricts to a homeomorphism

$$|K'| \xrightarrow{\cong} |K|,$$

and we can therefore sensibly call K' the **barycentric subdivision** of the simplicial complex K.

A natural question now arises: what relation is there between the simplicial homologies of K and K'? Their simplicial chain complexes are obviously not the same; in general, the chain complex for the subdivision K' has many more generators than that of K. But the polyhedra of these two complexes are the same, and it turns out that simplicial homology recognizes this fact. The result is best stated in terms of a concrete chain map

$$C_*^{\Delta}(K) \xrightarrow{S} C_*^{\Delta}(K'),$$

which can be defined by associating to each oriented n-simplex of K the n-chain of K' determined by its fundamental cycle. In the next lecture we will prove:

THEOREM 30.15. *The map $S_* : H_*^{\Delta}(K) \to H_*^{\Delta}(K')$ induced by the chain map described above is an isomorphism.*

30.5. Exercises.

Exercise 30.1. Prove Theorems 30.5 and 30.11 on the computation of $H_n^\Delta(K;\mathbb{Z}_2)$ and $H_n^\Delta(K;\mathbb{Z})$ for a closed and connected triangulated n-manifold $M \cong |K|$, in the second case with an oriented triangulation. Show moreover that if the triangulation does not admit any orientation, then $H_n^\Delta(K;\mathbb{Z}) = 0$.

Exercise 30.2. Prove Proposition 30.14, showing that the simplicial complex K with vertices in Δ^n defined via the barycentric subdivision algorithm actually defines a triangulation of Δ^n. *Hint: Argue inductively on n. Given any point $p \in \Delta^n$ distinct from the barycenter b_n, draw a straight line from b_n through p. What can you say about the point where this line exits Δ^n?*

31. Chain homotopy and simplicial approximation

I owe you an explanation of why Theorem 30.15 is true, but in this lecture I also want to sketch a deep application of this theorem, showing that the isomorphism class of the oriented simplicial homology $H_*^\Delta(K)$ of a finite simplicial complex K depends only on its polyhedron $|K|$. In other words, simplicial homology is a *topological* invariant, not just an invariant of abstract simplicial complexes. For concreteness, we shall work with the oriented rather than the ordered simplicial chain complex, which is not a loss of generality since we will also show in the next lecture that $H_*^o(K) \cong H_*^\Delta(K)$. A few tricky details will be omitted, and we will make up for this later in the semester by deriving a second proof of the topological invariance of $H_*^\Delta(K)$ from cellular homology. But several of the ideas discussed in this lecture will also be useful for other purposes, when we develop singular homology and study its applications.

In the early history of homology theory, it was widely believed that the topological invariance of simplicial homology should be deduced from Theorem 30.15 in combination with a result called the *Hauptvermutung*, which conjectured that any two triangulations of the same polyhedron could be made identical up to homeomorphism after sufficiently many iterations of the barycentric subdivision algorithm. At some point, the invariance of simplicial homology was proven by other means, and the Hauptvermutung remained an open question until it was, ironically, shown to be *false* in the 1960's. Theorem 30.15 can be viewed nonetheless as an important ingredient in a proof that $H_*^\Delta(K)$ depends only on $|K|$.

To state the main result properly, let

$$\mathsf{Cpct}^\Delta \subset \mathsf{Top}$$

denote the subcategory whose objects consist of all compact polyhedra, with arbitrary continuous maps as morphisms. We should clarify: a compact topological space X is an object of Cpct^Δ if and only if it is homeomorphic to the polyhedron $|K|$ of some finite simplicial complex, but the actual complex K and homeomorphism $X \cong |K|$ are *not* considered to be part of the data defining an object of Cpct^Δ. In general, a polyhedron has infinitely many distinct choices of possible triangulations, and without choosing specific triangulations, there is no canonical way to define what it means for a map between two polyhedra to be simplicial. This is one of a few reasons why we allow *all* continuous maps as morphisms in Cpct^Δ, rather than just simplicial maps.

Theorem 31.1. *There exists for each integer $n \geqslant 0$ and each R-module G a functor*

$$H_n^\Delta = H_n^\Delta(\cdot\,;G) : \mathsf{Cpct}^\Delta \to R\text{-}\mathsf{Mod}$$

that assigns to each compact polyhedron X the simplicial homology $H_n^\Delta(K;G)$ of some finite simplicial complex K whose polyhedron $|K|$ is homeomorphic to X.

Since homeomorphisms are isomorphisms in the category Cpct^Δ, this result implies:

Corollary 31.2. *If K and K' are two finite simplicial complexes with homeomorphic polyhedra $|K| \cong |K'|$, then their simplicial homologies $H_n^\Delta(K)$ and $H_n^\Delta(K')$ are isomorphic.* ☐

Notice what Theorem 31.1 does not say: we are not claiming that the functor $H_n^\Delta : \mathsf{Cpct}^\Delta \to R\text{-}\mathsf{Mod}$ is unique or canonical, and in fact, some arbitrary choices will need to be made in order to define it at the end of this lecture. The need for choices, however, does not detract from the power of the theorem: the mere fact that H_n^Δ is a functor on the category Cpct^Δ, whose morphisms are *arbitrary* continuous maps, is enough to deduce useful consequences such as Corollary 31.2.

31.1. The homotopy question. In preparation for proving Theorem 30.15, let us consider a slightly different question about the functoriality of simplicial homology. Suppose $f, g : K \to L$ are two simplicial maps between simplicial complexes such that the induced continuous maps of polyhedra $|K| \to |L|$ are homotopic. Does it follow that the induced homomorphisms

$$f_*, g_* : H_*^\Delta(K) \to H_*^\Delta(L)$$

are identical? We've seen that bordism theory has a homotopy invariance property of this type, and a similar property is also incorporated into the Eilenberg-Steenrod axioms for homology theories.

The assumption in the present context is that there exists a continuous map

$$I \times |K| \xrightarrow{h} |L|$$

with $h(0, \cdot) = f$ and $h(1, \cdot) = g$. In order to make something useful out of this in *simplicial* homology, it would seem natural to impose an extra condition and require h to be a *simplicial* map, but here we encounter an obstacle: it is not obvious whether $I \times |K|$ has a natural triangulation, which would be needed in order for the notion of a simplicial map to make sense. The polyhedron $|K|$ is a union of simplices $|\sigma| \cong \Delta^n$ of various dimensions $n \geqslant 0$, and this decomposes $I \times |K|$ into "prism-shaped" subsets of the form

$$I \times \Delta^n \cong \Delta^1 \times \Delta^n.$$

If we can find a sufficiently natural way of triangulating $\Delta^1 \times \Delta^n$, we will obtain from this a triangulation of $I \times |K|$ and thus be able to speak of *simplicial* homotopies $h : I \times |K| \to |L|$ between f and g.

31.2. Triangulating products of simplices. Let us frame the question a bit more generally: Is there a natural way to triangulate $\Delta^m \times \Delta^n$ for every pair of integers $m, n \geqslant 0$? This product of simplices is a manifold of dimension $m + n$ with boundary

$$\begin{aligned} \partial(\Delta^m \times \Delta^n) &= (\partial\Delta^m \times \Delta^n) \cup (\Delta^m \times \partial\Delta^n) \\ &= \left(\bigcup_{k=0}^m \partial_{(k)}\Delta^m \times \Delta^n \right) \cup \left(\bigcup_{k=0}^n \Delta^m \times \partial_{(k)}\Delta^n \right). \end{aligned}$$

Notice that each term in the union on the second line is canonically homeomorphic to a product of the form $\Delta^k \times \Delta^\ell$ for $k \leqslant m$ and $\ell \leqslant n$ with $k + \ell = m + n - 1$. This suggests an inductive condition that would be natural to require on our triangulations: if we assume that suitable triangulations of $\Delta^k \times \Delta^\ell$ have already been constructed for all $k + \ell < m + n$, then we would like our triangulation of $\Delta^m \times \Delta^n$ to reproduce these triangulations when restricted to its smooth boundary faces. We shall now describe a direct construction that produces this result.

Denote the standard basis of $\mathbb{R}^{m+n+2} = \mathbb{R}^{m+1} \times \mathbb{R}^{n+1}$ by

$$(e_0, 0), \ldots, (e_m, 0), (0, f_0), \ldots, (0, f_n) \in \mathbb{R}^{m+1} \times \mathbb{R}^{n+1},$$

so we can regard $e_0, \dots, e_m$ as the vertices of Δ^m and $f_0, \dots, f_n$ as the vertices of Δ^n. The triangulation of $\Delta^m \times \Delta^n$ we construct has vertex set

$$V := \big\{ (e_i, f_j) \in \Delta^m \times \Delta^n \bigm| i \in \{0, \dots, m\} \text{ and } j \in \{0, \dots, n\} \big\},$$

and its k-simplices for $k = 0, \dots, m+n$ will be the convex hulls of certain $(k+1)$-tuples of these vertices; in this way, the triangulation $\Delta^m \times \Delta^n \cong |K|$ will be uniquely determined once we have specified a suitable abstract simplicial complex $K = (V, S)$. To specify which subsets should be the vertices of a simplex in K, endow the set $\{0, \dots, m\} \times \{0, \dots, n\}$ with the total order such that $(i, j) \leqslant (i', j')$ if and only if $i \leqslant i'$ and $j \leqslant j'$, so *strict* inequality $(i, j) < (i', j')$ means additionally that $i < i'$ or $j < j'$. For $k = 0, \dots, m+n$, the k-simplices σ of K are then defined as

$$\sigma = \{(e_{i_0}, f_{j_0}), \dots, (e_{i_k}, f_{j_k})\} \subset \Delta^m \times \Delta^n,$$

for all possible strictly increasing sequences

$$(i_0, j_0) < \dots < (i_k, j_k) \in \{0, \dots, m\} \times \{0, \dots, n\}.$$

By this definition, we observe that the $(m+n)$-simplices all correspond to sequences $(i_0, j_0) < \dots < (i_{m+n}, j_{m+n})$ that begin with $(i_0, j_0) = (0, 0)$ and end with $(i_{m+n}, j_{m+n}) = (m, n)$, thus all of them contain the two specific vertices (e_0, f_0) and (e_m, f_n). Boundary faces σ of these $(m+n)$-simplices come in three types, corresponding to sequences $(i_0, j_0) < \dots < (i_{m+n-1}, j_{m+n-1})$ that satisfy the following conditions:

1. The sequence $j_0, \dots, j_{m+n-1}$ takes every value in $\{0, \dots, n\}$ but $i_0, \dots, i_{m+n-1}$ misses exactly one value $i \in \{0, \dots, m\}$.
2. The sequence $i_0, \dots, i_{m+n-1}$ takes every value in $\{0, \dots, m\}$ but $j_0, \dots, j_{m+n-1}$ misses exactly one value $j \in \{0, \dots, n\}$.
3. There are two consecutive terms of the form (i, j), $(i+1, j+1)$.

In the first two cases, the $m+n$ vertices of σ all lie in one of the convex sets

$$\partial_{(i)} \Delta^m \times \Delta^n \qquad \text{or} \qquad \Delta^m \times \partial_{(j)} \Delta^n.$$

As observed above, the union of these sets for all $i = 0, \dots, m$ and $j = 0, \dots, n$ is $\partial(\Delta^m \times \Delta^n)$, and these boundary faces thus determine an $(m+n-1)$-dimensional subcomplex $K' \subset K$ in which the convex hull of the vertices of each simplex is contained in $\partial(\Delta^m \times \Delta^n)$. It is easy to check that *all* simplices of K with convex hull contained in $\partial(\Delta^m \times \Delta^n)$ are of this form, because for any two points $p, q \in \partial(\Delta^m \times \Delta^n)$ that do not both belong to the same one of the $m+n+2$ convex subsets mentioned above, the line segment from p to q passes through the interior of $\Delta^m \times \Delta^n$. In particular, boundary faces of the third type in the list above do not belong to the subcomplex K'.

Since the vertices $(e_i, f_j) \in V$ are all points in $\Delta^m \times \Delta^n \subset \mathbb{R}^{m+n+2}$ and the latter is a convex set, the unique linear map $\mathbb{R}^V \to \mathbb{R}^{m+n+2}$ sending $e_v \mapsto v$ for each $v \in V$ restricts to the polyhedron $|K| \subset \mathbb{R}^V$ as a map

$$|K| \to \Delta^m \times \Delta^n. \tag{31.1}$$

Exercise 31.1 shows that this map is a homeomorphism, and thus defines a triangulation of $\Delta^m \times \Delta^n$; moreover, restricting it to the subcomplex K' formed by vertices contained in $\partial(\Delta^m \times \Delta^n)$ gives a homeomorphism $|K'| \cong \partial(\Delta^m \times \Delta^n)$.

In order to define fundamental cycles $c_{\Delta^m \times \Delta^n}$ in $H^o_{m+n}(K; \mathbb{Z})$ and $H^\Delta_{m+n}(K; \mathbb{Z})$ from our triangulation, we need to endow it with an orientation and choose an admissible ordering for the simplicial complex K. The latter is easy, as the total order on $\{0, \dots, k\} \times \{0, \dots, \ell\}$ determines a total order on the set of all vertices of K. In order to define a suitable orientation, let $\mathbf{S}(m, n)$ denote the set of all strictly increasing sequences $(i_0, j_0) < \dots < (i_{m+n}, j_{m+n})$ of $m+n+1$

elements in $\{0, \ldots, m\} \times \{0, \ldots, n\}$, and write $\sigma_{\mathbf{s}}$ for the $(m+n)$-simplex of K determined by each $\mathbf{s} \in \mathbf{S}(m,n)$. Denote by $\mathbf{s}_0 \in \mathbf{S}(m,n)$ the specific sequence

$$(0,0) < (1,0) < \ldots < (m,0) < (m,1) < \ldots < (m,n),$$

and define the **parity** $|\mathbf{s}| \in \mathbb{Z}_2$ of any element $\mathbf{s} \in \mathbf{S}(m,n)$ to be the number of steps (modulo 2) required in order to transform $\mathbf{s}_0$ into $\mathbf{s}$ via operations that modify three consecutive terms of a sequence like so:

$$(i-1,j) < (i,j) < (i,j+1) \qquad \rightsquigarrow \qquad (i-1,j) < (i-1,j+1) < (i,j+1).$$

Lemma 31.3. *The parity $|\mathbf{s}| \in \mathbb{Z}_2$ of elements $\mathbf{s} \in \mathbf{S}(m,n)$ is independent of choices.*

Proof sketch. We can interpret $(-1)^{|\mathbf{s}|} \in \{1,-1\}$ as the sign of a permutation of $m+n$ elements, which include m copies of the letter R (for "right") and n copies of the letter U (for "up"). ☐

The chosen orientation and admissible ordering for K determine a fundamental cycle

$$c_{\Delta^m \times \Delta^n} \in C^{\Delta}_{m+n}(K;\mathbb{Z}) \quad \text{or} \quad C^{o}_{m+n}(K;\mathbb{Z}).$$

For each $m,n \geqslant 1$, the relation $\partial c_{\Delta^m \times \Delta^n} = c_{\partial(\Delta^m \times \Delta^n)}$ then becomes the following formula under the usual identification between boundary faces and simplices of one dimension lower:

$$\partial c_{\Delta^m \times \Delta^n} = \sum_{i=0}^{m} (-1)^i c_{\partial_{(i)}\Delta^m \times \Delta^n} + (-1)^m \sum_{j=0}^{n} (-1)^j c_{\Delta^m \times \partial_{(j)}\Delta^n}. \tag{31.2}$$

Remark 31.4. If we regard $\partial\Delta^m \times \Delta^n$ and $\Delta^m \times \partial\Delta^n$ as compact topological $(m+n-1)$-manifolds with matching boundary $\partial\Delta^m \times \partial\Delta^n$ and endow both with the obvious oriented triangulations and admissible orderings that they inherit from $\Delta^m \times \Delta^n$, the formula in (31.2) takes the slightly prettier form

$$\partial c_{\Delta^m \times \Delta^n} = c_{\partial\Delta^m \times \Delta^n} + (-1)^m c_{\Delta^m \times \partial\Delta^n}.$$

When we introduce the homological cross product later in this course, the singular homology version of this relation will take the form

$$\partial(c_{\Delta^m} \times c_{\Delta^n}) = \partial c_{\Delta^m} \times c_{\Delta^n} + (-1)^m c_{\Delta^m} \times \partial c_{\Delta^n},$$

which is written in terms of the obvious fundamental cycle $c_{\Delta^k} \in C_k(\Delta^k;\mathbb{Z})$ for the standard simplex of each dimension with its trivial triangulation, and a bilinear product operation

$$C_*(X) \otimes C_*(Y) \to C_*(X \times Y) : A \otimes B \mapsto A \times B$$

that relates the singular chain complexes of any two spaces X, Y and sends $C_m(X) \otimes C_n(Y)$ in general to $C_{m+n}(X \times Y)$. We will have plenty to say about this product later, but the detail I want to comment on right now is the sign $(-1)^m$ appearing on the right hand side of the formula. This is an instance of a general pattern known as the **Koszul sign convention**, which we will see many more examples of in this course. In a nutshell, the rule is that whenever objects carry natural gradings in either $\mathbb{Z}$ or $\mathbb{Z}_2$, exchanging the order of two objects with odd degree causes a sign change. In the present context, the "objects" to which this rule applies are not only the chains of certain degrees in Δ^m and Δ^n but also the operator ∂, which we regard as having degree -1 since it maps k-chains to $(k-1)$-chains for every k. This means that no sign change is necesary when writing $\partial c_{\Delta^m} \times c_{\Delta^n}$, since the three objects ∂, c_{Δ^m} and c_{Δ^n} appear here in the same order as on the left hand side, but writing $c_{\Delta^m} \times \partial c_{\Delta^n}$ exchanges the order of ∂ and c_{Δ^m}, and since ∂ has odd degree, a sign change must then result if and only if c_{Δ^m} also has odd degree, meaning m is odd. One could presumably state some general theorem in category-theoretic terms to explain why and in what contexts this particular way of dealing with signs gives the results we want, but

I personally would consider writing down that theorem to be more trouble than it is worth. If you haven't seen the Koszul convention before in one of the many other contexts (e.g. the exterior algebra of differential forms on smooth manifolds) where it naturally arises, then I think that you will in any case learn through experience during the remainder of this course why it is good and useful.

31.3. From simplicial homotopies to chain homotopies. Let us identify the unit interval I with the standard 1-simplex via the homeomorphism

$$I \xrightarrow{\cong} \Delta^1 \subset \mathbb{R}^2 : t \mapsto (1-t, t).$$

The oriented triangulation of $\Delta^1 \times \Delta^n$ constructed in the previous section for each $n \geqslant 0$ yields an oriented triangulation of $I \times \Delta^n$ whose restriction to the smooth faces of

$$\partial(I \times \Delta^n) = (\{1\} \times \Delta^n) \cup (\{0\} \times \Delta^n) \cup \left(\bigcup_{k=0}^{n} I \times \partial_{(k)}\Delta^n \right)$$

matches the trivial triangulation of Δ^n and the constructed trivialization of $I \times \Delta^{n-1}$.

Now consider again the polyhedron $|K|$ discussed in §31.1 above, and choose an admissible ordering for the underlying simplicial complex K. The ordering determines an identification of each n-simplex of $|K|$ for $n = 0, 1, 2, \ldots$ with the standard n-simplex, and applying the triangulation algorithm then gives a triangulation of $I \times |K|$, whose underlying simplicial complex we shall denote in the following by K_I. For this triangulation, the two inclusions

$$|K| \overset{\iota_j}{\hookrightarrow} I \times |K| : p \mapsto (j, p) \qquad \text{for } j = 0, 1$$

are both simplicial maps. Now suppose $\sigma = [v_0, \ldots, v_n]$ is an n-simplex of K, equipped with the orientation it inherits from the admissible ordering, and let $|\sigma| \subset |K|$ denote the corresponding subset homeomorphic to Δ^n in the polyhedron. Our triangulation determines an oriented triangulation of the $(n+1)$-dimensional manifold $I \times \Delta^n \cong I \times |\sigma| \subset I \times |K|$, thus a fundamental cycle

$$c_{I \times |\sigma|} \in C^{\Delta}_{n+1}(K_I; \mathbb{Z}),$$

for which the formula (31.2) specializes to this situation as

$$\partial c_{I \times |\sigma|} = (\iota_1)_* \sigma - (\iota_0)_* \sigma - c_{I \times \partial|\sigma|}.$$

Here $c_{I \times \partial|\sigma|}$ is an abbreviation for the signed sum of fundamental cycles of the induced triangulation of $I \times \Delta^{n-1}$ for each boundary face $\Delta^{n-1} \cong \partial_{(k)}\Delta^n$ of $|\sigma| \cong \Delta^n$.

Since $I \times |K|$ is now a polyhedron, we can sensibly impose an extra condition on the homotopy $h : I \times |K| \to |L|$, and require it to be a simplicial map, i.e. a **simplicial homotopy** between f and g. With this assumption in place, there is a unique homomorphism

$$C^{\Delta}_n(K) \xrightarrow{h_\#} C^{\Delta}_{n+1}(L)$$

defined for each $n \geqslant 0$ via linearity and the formula

$$h_\#(\sigma) := h_* c_{I \times |\sigma|}.$$

for oriented n-simplices $\sigma \in [v_0, \ldots, v_n]$ of K. Since $h_* : C_*(K_I) \to C_*(L)$ is a chain map, the formula above for $\partial c_{I \times |\sigma|}$ implies

$$\partial h_\#(\sigma) = g_* \sigma - f_* \sigma - h_\#(\partial \sigma),$$

so that $h_\#$ satisfies the so-called **chain homotopy relation**

$$\partial h_\# + h_\# \partial = g_* - f_*.$$

A brief algebraic digression is now in order.

Definition 31.5. Given two chain maps $f, g : (A_*, \partial^A) \to (B_*, \partial^B)$, a **chain homotopy** (*Kettenhomotopie*) from f to g is a homomorphism $h : A_* \to B_*$ that satisfies $h(A_n) \subset B_{n+1}$ for each $n \in \mathbb{Z}$ and the chain homotopy relation

$$\partial^B h + h\partial^A = g - f.$$

We say that f and g are **chain homotopic** if there exists a chain homotopy from f to g.

One easily checks that the notion of chain homotopy defines an equivalence relation between chain maps, and moreover, if f_0 and f_1 are chain homotopic and have well-defined compositions of chain maps $f_j \circ g$, then $f_0 \circ g$ and $f_1 \circ g$ are also chain homotopic; a similar statement applies to compositions of the form $g \circ f_j$. The upshot is that there is a well-defined **homotopy category of chain complexes**

$$\mathsf{hCh}(R\text{-}\mathsf{Mod}) \qquad \text{abbreviated as} \qquad \mathsf{hCh},$$

in which the objects are chain complexes of R-modules and the morphisms are chain homotopy classes of chain maps. An isomorphism in the category hCh is called a **chain homotopy equivalence** (*Kettenhomotopieäquivalenz*), so a chain map $f : A_* \to B_*$ is a chain homotopy equivalence if and only if it admits a **chain homotopy inverse** $g : B_* \to A_*$, meaning a chain map such that the compositions $g \circ f$ and $f \circ g$ are each chain homotopic to identity maps.

There are two convincing reasons why the category hCh is important to define: the first is that simplicial homotopies between simplicial maps give rise to chain homotopies between the induced chain maps, as shown above—and we will see later that chain homotopies in the singular chain complex similarly arise from arbitrary homotopies between continuous maps. The second reason is the following easy result, which tells us that the algebraic homology functors $H_n : \mathsf{Ch} \to R\text{-}\mathsf{Mod}$ descend to the homotopy category as functors $\mathsf{hCh} \to R\text{-}\mathsf{Mod}$.

Proposition 31.6. *If $f, g : A_* \to B_*$ are chain homotopic chain maps, then for each $n \in \mathbb{Z}$, the homomorphisms $f_*, g_* : H_n(A_*) \to H_n(B_*)$ they induce on homology are identical.*

Proof. Given $[a] \in H_n(A_*)$, the representative $a \in A_n$ is a cycle, so the chain homotopy relation gives $g(a) - f(a) = \partial^B h(a) + h\partial^A a = \partial^B h(a)$, implying $[f(a)] = [g(a)] \in H_n(B_*)$. ☐

Putting all of this together implies:

Corollary 31.7. *If $f, g : |K| \to |L|$ are simplicial maps related by a simplicial homotopy, then for each $n \in \mathbb{Z}$, the induced maps $f_*, g_* : H_n^\Delta(K) \to H_n^\Delta(L)$ are identical.* ☐

We have used oriented simplicial homology in this discussion for the sake of concreteness, but the discussion also makes sense for ordered simplicial homology.

31.4. Subdivision defines a chain homotopy equivalence. Now that the notion of a chain homotopy equivalence has been defined, we can explain the real reason behind Theorem 30.15. Assume K is a simplicial complex and K' is the complex defined from K by barycentric subdivision, giving rise to the chain map

$$S : C_*^\Delta(K) \to C_*^\Delta(K')$$

described in the previous lecture. A special class of chain maps in the other direction can be defined as follows. By definition, every vertex v in the complex K' is the barycenter of a particular simplex σ_v in the polyhedron $|K|$; note that this includes the vertices of K' that are also vertices of K, since the latter are also 0-simplices of K. For each vertex v of K', let $w(v)$ denote an arbitrary choice of a vertex of the simplex σ_v in K that has v as its barycenter. One can check that this defines a simplicial map

$$K' \xrightarrow{\pi} K : v \mapsto w_v,$$

and we call it a **projection** since it necessarily sends each vertex of K' that is also a vertex of K to itself. The following result now implies Theorem 30.15:

THEOREM 31.8. *For any choice of projection* $\pi : K' \to K$*, the induced chain map* $\pi_* : C_*^\Delta(K') \to C_*^\Delta(K)$ *is a chain homotopy inverse of* $S : C_*^\Delta(K) \to C_*^\Delta(K')$*, implying in particular that the latter is a chain homotopy equivalence.*

A complete proof of this theorem can be found e.g. in [**ES52**, Theorem VI.7.1]; here we shall content ourselves with a brief sketch. One can verify directly that $\pi_* S : C_*^\Delta(K) \to C_*^\Delta(K)$ is the identity map. It then remains to show that

$$S\pi_* : C_*^\Delta(K') \to C_*^\Delta(K')$$

is chain homotopic to the identity. The details of this chain homotopy would require too much of a digression, but there is a geometric construction in the background that is worth understanding: in a different context, the same construction will later give us a relatively straightforward construction of a chain homotopy for the natural subdivision operator on singular homology.

The construction is yet another triangulation of the prism $I \times \Delta^n$, one that interpolates between the trivial triangulation of $\{0\} \times \Delta^n$ and the barycentric subdivision of $\{1\} \times \Delta^n$. Like the other explicit triangulations we've discussed, it decomposes $I \times \Delta^n$ into convex regions determined by sets of vertices in general position, and it admits an inductive description: for $n = 0$, one takes the obvious triangulation of $I \times \Delta^0 \cong I$ with a single 1-simplex. Assuming that a suitable triangulation of the n-manifold $I \times \Delta^{n-1}$ for some $n \geqslant 1$ has already been constructed, the $(n+1)$-simplices of our triangulation of $I \times \Delta^n$ then come in two types:

- One whose vertices are $(0, e_0), \ldots, (0, e_n)$ and $(1, b_n)$, where $e_0, \ldots, e_n$ are the standard basis vectors of $\mathbb{R}^{n+1}$ (i.e. the vertices of Δ^n) and $b_n \in \Delta^n$ is the barycenter.
- For each $k = 0, \ldots, n$ and each n-simplex $\{v_0, \ldots, v_n\}$ in the triangulation of $I \times \partial_{(k)}\Delta^n \cong I \times \Delta^{n-1}$, one with vertices $v_0, \ldots, v_n$ and $(1, b_n)$.

Exercise 31.2 implies that this defines an oriented triangulation of $I \times \Delta^n$.

Intuitively, the three pieces of the boundary

$$\partial(I \times \Delta^n) = (\{1\} \times \Delta^n) \cup (\{0\} \times \Delta^n) \cup (I \times \partial\Delta^n)$$

with their induced triangulations now correspond to the three terms on the right hand side of a chain homotopy relation

$$\partial h_\# = S\pi_* - \mathbb{1} - h_\# \partial$$

for some chain homotopy $h_\# : C_n^\Delta(K') \to C_{n+1}^\Delta(K')$. In the simplicial context, it is not so straightforward to make this intuition precise, but we will return to this subject in the near future in the context of singular homology, where the definition of the corresponding chain homotopy is more straightforward.

31.5. Simplicial approximation.

The last major ingredient needed for a proof of Theorem 31.1 is a result that relates the categories Cpct^Δ and Simp:

THEOREM 31.9 (simplicial approximation). *If* $X \cong |K|$ *and* $Y \cong |L|$ *are compact polyhedra and* $f : X \to Y$ *is a continuous map, then after finitely-many iterations of barycentric subdivision to replace the triangulation of* X *with a finer triangulation* $X \cong |K'|$*,* f *is homotopic to a map* $g : X \to Y$ *that arises from a simplicial map* $K' \to L$*. Moreover, for every* $x \in X$*,* $g(x)$ *is contained in the smallest simplex of* Y *containing* $f(x)$*.*

We might have naively hoped for the theorem to state that every continuous map between polyhedra with fixed choices of triangulations is homotopic to a simplicial map—but there are easy counterexamples to that statement. For instance, every non-surjective map $S^1 \to S^1$ has its image in a contractible space $S^1\backslash\{*\} \cong \mathbb{R}$ and is thus homotopic to a constant, implying that every map $S^1 \to S^1$ homotopic to the identity is surjective. But if we choose two triangulations $S^1 \cong |K|$ and $S^1 \cong |L|$ such that L has strictly more vertices than K, then no simplicial map $K \to L$ can be surjective, and the identity $S^1 \to S^1$ therefore cannot be homotopic to any simplicial map with respect to these particular triangulations. Of course, this problem goes away if we are also allowed to replace K with a triangulation that has arbitrarily many vertices, e.g. by iterated barycentric subdivision.

We refer to [**Hat02**, §2.C] for a detailed proof of Theorem 31.9, but the following explains the basic idea.

Sketch of the proof of Theorem 31.9. For each vertex $v \in X$, define the so-called **open star** of v as the open neighborhood

$$\operatorname{st} v \subset X$$

of v formed by the union of the interiors of all simplices in X that have v as a vertex. Figure 17 shows the open stars of two neighboring vertices in a 2-dimensional polyhedron; notice that their intersection contains the interior of the 1-simplex bounded by these two vertices (cf. Exercise 31.3). The collection of all open stars of vertices defines an open covering of any polyhedron. Now given $f : X \to Y$ continuous, after subdividing the triangulation of X enough times, we can assume that for every vertex $v \in X$ there exists a vertex $w_v \in Y$ such that (see Figure 17 again)

$$\operatorname{st} v \subset f^{-1}(\operatorname{st} w_v).$$

Having associated to each $v \in X$ some $w_v \in Y$ with this property, there is a unique simplicial map $g : X \to Y$ that satisfies $g(v) = w_v$: indeed, for every simplex $\{v_0, \dots, v_n\}$ of X, Exercise 31.3 implies that the set $\{w_{v_0}, \dots, w_{v_n}\}$ is also a simplex of Y. One can now check that g is indeed an "approximation" of f in the sense that $g(x)$ is contained in the smallest simplex of Y containing $f(x)$ for every $x \in X$. In light of this, a homotopy $h : I \times X \to Y$ from f to g can be defined by choosing $h(\cdot, x) : I \to Y$ for every $x \in X$ to be the linear path from $f(x)$ to $g(x)$ in the smallest simplex containing $f(x)$. ☐

31.6. Simplicial homology as a topological invariant. Here is a sketch of a proof of Theorem 31.1.

By the axiom of choice, we can associate to every compact polyhedron $X \in \mathsf{Cpct}^\Delta$ a specific choice of finite simplicial complex K_X and triangulation $X \cong |K_X|$; having done this, define

$$H_n^\Delta(X) := H_n^\Delta(K_X).$$

For each continuous map $f : X \to Y$ between compact polyhedra, we can apply the simplicial approximation theorem to find a sufficiently fine subdivision K_X' of K_X and a simplicial map $g : K_X' \to K_Y$ for which the associated continuous map $g : X \to Y$ is homotopic to f. Writing $S_* : H_n^\Delta(K_X) \to H_n^\Delta(K_X')$ for the isomorphism defined via iterated barycentric subdivision, the homomorphism $f_* : H_n^\Delta(X) \to H_n^\Delta(Y)$ induced by f can then be defined by

$$\begin{tikzcd} H_n^\Delta(X) = H_n^\Delta(K_X) \arrow[r, "S_*", "\cong"'] \arrow[rr, bend left, "f_*"] & H_n^\Delta(K_X') \arrow[r, "g_*"] & H_n^\Delta(K_Y) = H_n^\Delta(Y) \end{tikzcd}.$$

The result of §31.3 on simplicial homotopies can be used in showing that the map $f_* : H_n^\Delta(X) \to H_n^\Delta(Y)$ defined in this way is independent of choices. Putting all this together produces a functor $H_n^\Delta : \mathsf{Cpct}^\Delta \to R\text{-}\mathsf{Mod}$.

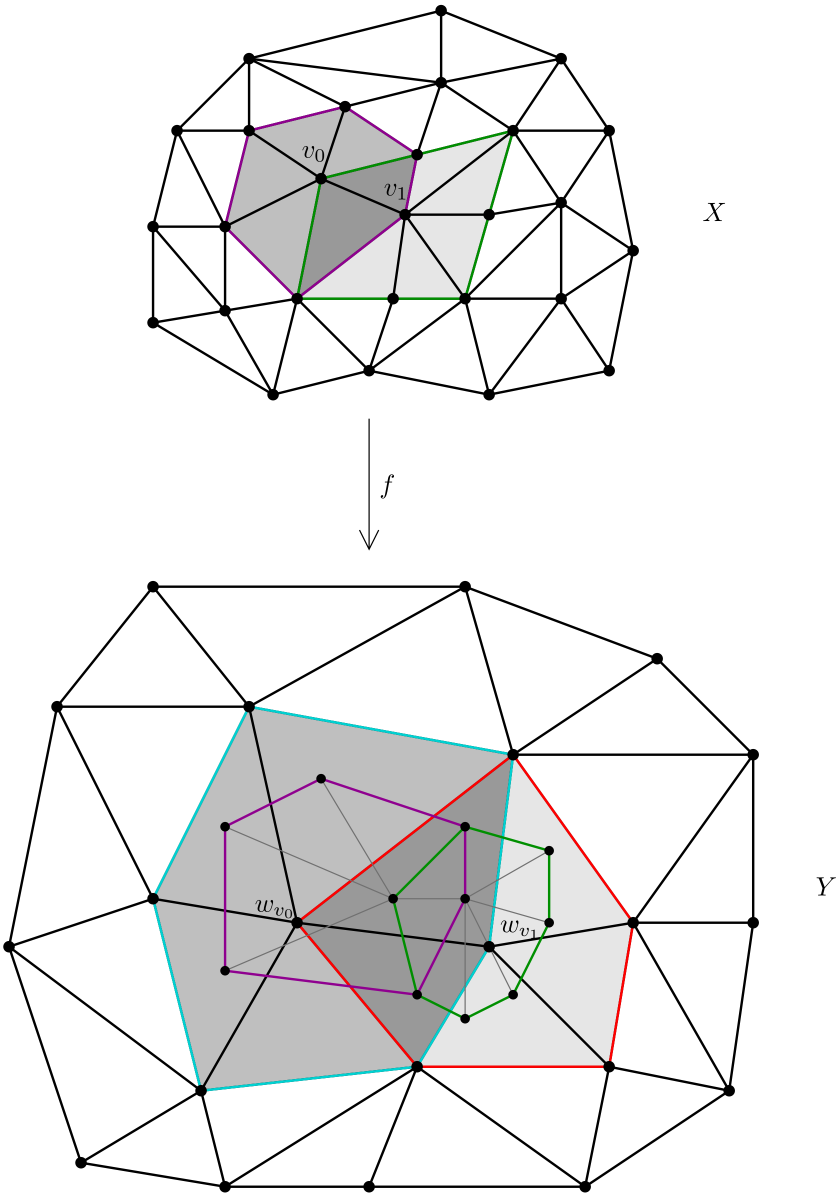


FIGURE 17. A map $f : X \to Y$ between two polyhedra, with vertices $v_0, v_1 \in X$ and $w_{v_0}, w_{v_1} \in Y$ chosen such that f maps the open star of v_i into the open star of w_{v_i} for $i = 0, 1$. The prescription in the proof of Theorem 31.9 will then produce a simplicial map $g : X \to Y$ sending $v_i \mapsto w_{v_i}$ for $i = 0, 1$, so the 1-simplex in X bounded by v_0 and v_1 is sent to the 1-simplex in Y bounded by w_{v_0} and w_{v_1}.

31.7. Exercises.

EXERCISE 31.1. Prove that the map $|K| \to \Delta^m \times \Delta^n$ described in (31.1) is a homeomorphism. *Hint: This is probably not the only possible approach, but here an inductive argument as in Exercise 30.2 is also possible. Use the fact that certain points are contained in all the n-simplices.*

Exercise 31.2. Let L denote the $(n+1)$-dimensional abstract simplicial complex formed by the sets of vertices in $I \times \Delta^n$ described in §31.4, and define $L' \subset L$ to be the subcomplex of simplices whose vertices have convex hulls lying in $\partial(I \times \Delta^n)$.

(a) Carry out the analogue of Exercise 30.2 to show that L defines a triangulation of $I \times \Delta^n$ for which the subcomplex L' triangulates $\partial(I \times \Delta^n)$.
Hint: To show that every point $p \in I \times \Delta^n$ lies in one of the $(n+1)$-simplices described, draw a line from $(1, b_n)$ through p and see where it exits through $\partial(I \times \Delta^n)$.

(b) Describe an inductive algorithm to produce suitable admissible orderings and orientations for this triangulation of $I \times \Delta^n$ for each $n \geqslant 0$.

Exercise 31.3. Given vertices $v_0, \dots, v_k$ in a polyhedron X, show that $\bigcap_{i=0}^k \operatorname{st} v_i \neq \varnothing$ if and only if X contains a simplex whose vertices are $v_0, \dots, v_k$.

32. Acyclic models and relative homology

I want to tie up a few loose ends regarding simplicial homology before we move on to singular homology in the next lecture. One important topic is the reason why the ordered simplicial homology $H_*^o(K)$ and its oriented counterpart $H_*^\Delta(K)$ are isomorphic: we will prove this using the method of acyclic models, which will also be quite useful in our later discussion of products in singular homology and cohomology. We also take this opportunity to introduce *relative* simplicial homology, and explain the general algebraic mechanism that leads to long exact sequences of homology groups.

Several results in this lecture will apply equally well to the ordered and oriented versions of simplicial homology, and the following notational convention will allow us to talk about both at the same time:

$$H_*^\bullet := H_*^o \text{ or } H_*^\Delta,$$
$$C_*^\bullet := C_*^o \text{ or } C_*^\Delta.$$

32.1. Reduced simplicial homology. We discussed the reduced version $\widetilde{h}_*$ of an axiomatic homology theory h_* in Lecture 28. A reduced version of simplicial homology can be defined analogously, after observing that the one-point space $\{*\}$ is a polyhedron, whose underlying simplicial complex consists only of a single vertex. We shall also denote this one-point simplicial complex by $\{*\}$, and let

$$K \xrightarrow{\epsilon} \{*\}$$

denote the unique simplicial map from any given simplicial complex K to the one-point complex. The reduced (ordered or oriented) simplicial homology is then defined by

$$\widetilde{H}_n^\bullet(K) = \widetilde{H}_n^\bullet(K; G) := \ker\left(H_n^\bullet(K; G) \xrightarrow{\epsilon_*} H_n^\bullet(\{*\}; G)\right).$$

As with axiomatic homology, we can always choose a right-inverse of $\epsilon : K \to \{*\}$, which in this context must be a simplicial map $\{*\} \hookrightarrow K$, and the induced homomorphism on homology gives rise to a splitting of the short exact sequence

$$0 \to \widetilde{H}_n^\bullet(K) \hookrightarrow H_n^\bullet(K) \overset{\epsilon_*}{\to} H_n^\bullet(\{*\}) \to 0,$$

and thus an isomorphism $H_n^\bullet(K) \cong \widetilde{H}_n^\bullet(K) \oplus H_n^\bullet(\{*\})$. We recall from Exercise 29.4 that $H_n^\bullet(\{*\}; G)$ is trivial for $n \neq 0$ and is naturally isomorphic to the coefficient group G for $n = 0$, so the result is

$$H_n^\bullet(K; G) \cong \begin{cases} \widetilde{H}_n^\bullet(K; G) \oplus G & \text{for } n = 0, \\ \widetilde{H}_n^\bullet(K; G) & \text{for } n \neq 0. \end{cases}$$

Working through Exercise 29.4 also leads to the following observation: In both versions of the simplicial chain complex for $\{*\}$, the boundary map $\partial_1 : C_1^\bullet(\{*\}) \to C_0^\bullet(\{*\})$ at degree 1 is trivial. Indeed, this is immediate in the oriented chain complex because $C_1^\Delta(\{*\})$ is trivial due to the lack of 1-simplices, while in the ordered chain complex, $C_1^o(\{*\})$ has only a single generator $(*, *)$ determined by the unique vertex $* \in \{*\}$, which satisfies

$$\partial(*, *) = (*) - (*) = 0.$$

Since $\epsilon_* : C_*^\bullet(K) \to C_*^\bullet(\{*\})$ is a chain map, the relation $\epsilon_*\partial = \partial\epsilon_*$ then implies that the composition of $\partial_1 : C_1^\bullet(K) \to C_0^\bullet(K)$ with the so-called **augmentation** $\epsilon_* : C_0^\bullet(K) \to C_0^\bullet(\{*\}) = G$ is trivial, leading to the so-called **augmented chain complex**

$$\ldots \longrightarrow C_2^\bullet(K;G) \xrightarrow{\partial_2} C_1^\bullet(K;G) \xrightarrow{\partial_1} C_0^\bullet(K;G) \xrightarrow{\epsilon_*} G \longrightarrow 0 \longrightarrow 0 \longrightarrow \ldots,$$

in which we use the natural isomorphism $C_0^\bullet(\{*\};G) \cong G$ to replace $C_0^\bullet(\{*\};G)$ by the coefficient group G, and the map $\epsilon_* : C_0^\bullet(K;G) \to G$ can then be expressed via the direct formula

$$\epsilon_*\left(\sum_i a_i\sigma_i\right) = \sum_i a_i$$

for any finite linear combination of generators σ_i with coefficients $a_i \in G$. We shall denote the augmented chain complex by $\widetilde{C}_*^\bullet(K) = \widetilde{C}_*^\bullet(K;G)$, with chain groups

$$\widetilde{C}_n^\bullet(K;G) := \begin{cases} C_n^\bullet(K;G) & \text{for } n \neq -1, \\ G & \text{for } n = -1, \end{cases}$$

and boundary map $\partial : \widetilde{C}_*^\bullet(K) \to \widetilde{C}_*^\bullet(K)$ matching that of the usual chain complex $C_*^\bullet(K)$ except at degree 0, where it is defined to be the augmentation $\epsilon_* : C_0^\bullet(K;G) \to G$. The following result is a near immediate consequence of the definitions.

PROPOSITION 32.1. *There is a natural isomorphism*

$$H_*\big(\widetilde{C}_*^\bullet(K;G)\big) \xrightarrow{\cong} \widetilde{H}_*^\bullet(K;G)$$

that takes the form $[c] \mapsto [c]$ *for cycles* $c \in \widetilde{C}_n^\bullet(K;G)$ *of degree* $n \geqslant 0$. □

32.2. The cone of a simplicial complex. The point of defining reduced simplicial homology is to have a version of simplicial homology that vanishes in *all* degrees for certain contractible polyhedra that arise in applications. Here is a popular class of examples.

DEFINITION 32.2. The **cone** of a simplicial complex $K = (V, S)$ is the simplicial complex $CK = (CV, CS)$ with vertices

$$CV := V \cup \{*\}$$

and simplices

$$CS := S \cup \big\{\{v_0, \ldots, v_n, *\} \;\big|\; \{v_0, \ldots, v_n\} \in S\big\},$$

where $*$ denotes an *extra* vertex that is assumed to be not an element of the original vertex set V.

The polyhedron $|CK|$ of a cone complex CK has an obvious identification with the topological cone $C|K|$ of the original polyhedron $|K|$, in which the summit of the cone corresponds to the extra vertex $* \in CV$.

DEFINITION 32.3. A chain complex A_* is called **chain contractible** if the identity map $\mathbb{1} : A_* \to A_*$ is chain homotopic to the trivial chain map $0 : A_* \to A_*$.

If A_* is chain contractible, then looking at induced maps $H_*(A_*) \to H_*(A_*)$, we find that the identity map and the zero map on $H_*(A_*)$ must be identical, which is only possible if $H_*(A_*) = 0$. A chain complex with the latter property is said to be **acyclic**, in other words, A_* has no cycles other than those which are trivial in the sense of being boundaries. Chain contractible complexes are thus acyclic; one can view this as an algebraic counterpart to the topological fact that contractible spaces have trivial reduced homology according to the axioms.

Lemma 32.4. *For any simplicial complex K, the augmented simplicial chain complex $\widetilde{C}^{\bullet}_*(CK;\mathbb{Z})$ of its cone is chain contractible.*

Proof. Let us write down a proof for the ordered chain complex, from which a proof for the oriented complex can be obtained just be changing round brackets into square brackets. For each $n \geqslant 0$, we can specify a homomorphism $h_\# : \widetilde{C}^o_n(CK;\mathbb{Z}) \to \widetilde{C}^o_{n+1}(CK;\mathbb{Z})$ by saying how it is defined on an arbitrary generator $(v_0,\dots,v_n) \in C^o_n(CK;\mathbb{Z}) = \widetilde{C}^o_n(CK;\mathbb{Z})$, so we define

$$\widetilde{C}^o_n(CK;\mathbb{Z}) \xrightarrow{h_\#} \widetilde{C}^o_{n+1}(CK;\mathbb{Z}) : (v_0,\dots,v_n) \mapsto (*,v_0,\dots,v_n),$$

and we extend it to $n = -1$ by specifying its value on the generator $1 \in \mathbb{Z} = \widetilde{C}^o_{-1}(CK;\mathbb{Z})$, namely

$$\widetilde{C}^o_{-1}(CK;\mathbb{Z}) \xrightarrow{h_\#} \widetilde{C}^o_0(CK;\mathbb{Z}) : 1 \mapsto (*).$$

We then have

$$\partial h_\#(v_0,\dots,v_n) = (v_0,\dots,v_n) - h_\#\partial(v_0,\dots,v_n) \qquad \text{and} \qquad \partial h_\#(1) = \epsilon_*(*) = 1 = 1 - h_\#\partial(1)$$

since $\widetilde{C}^o_{-2}(CK;\mathbb{Z})$ is trivial by definition and thus $\partial(1) = 0$. This establishes the chain homotopy relation $\partial h_\# + h_\#\partial = \mathbb{1} = \mathbb{1} - 0$. ☐

Example 32.5. For some $n \geqslant 1$, suppose K is a simplicial complex containing only a single n-simplex and all its faces, so $|K| \cong \Delta^n$. Then K can be identified with the cone of a complex K' with $|K'| \cong \Delta^{n-1}$, and the lemma above therefore implies that $\widetilde{C}_*(K;\mathbb{Z})$ is chain contractible.

32.3. Natural chain homotopy equivalences. Recall that for any simplicial complex K and any choice of coefficients, the quotient projection $(v_0,\dots,v_n) \mapsto [v_0,\dots,v_n]$ determines a natural chain map

$$C^o_*(K) \xrightarrow{\Psi_K} C^\Delta_*(K).$$

Here the word *natural* carries a precise meaning that will be important to clarify: it means that for any other simplicial complex L with a simplicial map $f : L \to K$, the diagram

$$\begin{tikzcd}
C^o_*(L) \arrow[r, "\Psi_K"] \arrow[d, "f_*"] & C^\Delta_*(L) \arrow[d, "f_*"] \\
C^o_*(K) \arrow[r, "\Psi_L"] & C^\Delta_*(K)
\end{tikzcd}$$

commutes. As a special case, suppose $L \subset K$ is a subcomplex and $f : L \hookrightarrow K$ is the inclusion map: the chain maps $f_* : C^o_*(L) \to C^o_*(K)$ and $f_* : C^\Delta_*(L) \to C^\Delta_*(K)$ are then likewise inclusions of subcomplexes, and naturality then implies firstly that Ψ_K sends the subcomplex $C^o_*(L) \subset C^o_*(K)$ into the subcomplex $C^\Delta_*(L) \subset C^\Delta_*(K)$, and secondly that Ψ_L is simply the restriction of Ψ_K to $C^o_*(L)$.

With this observation as motivation, let us say more generally that for a specific simplicial complex K, a chain map $\Psi : C^o_*(K) \to C^\Delta_*(K)$ is **natural** if for every subcomplex $L \subset K$, Ψ sends $C^o_*(L)$ into $C^\Delta_*(L)$. In the same manner, one can define the notion of a *natural* chain map in the other direction $C^\Delta_*(K) \to C^o_*(K)$, or between each of $C^\Delta_*(K)$ or $C^o_*(K)$ and itself. Such chain maps can always be interpreted as natural transformations between two functors to the category of chain complexes, defined on a category that has subcomplexes of K as objects and inclusion maps as morphisms.

The following result explains why the natural homomorphism $H^o_*(K) \to H^\Delta_*(K)$ induced by the chain map Ψ_K is always an isomorphism, thus making the ordered and oriented versions of simplicial homology interchangeable in practice.

THEOREM 32.6. *For every simplicial complex K, the natural chain map $\Psi_K : C^o_*(K) \to C^\Delta_*(K)$ is a chain homotopy equivalence.*

The theorem will follow from three lemmas, each of which should be understood to hold for an arbitrary simplicial complex K:

LEMMA 32.7. *There exists a natural chain map $\Phi : C^\Delta_*(K;\mathbb{Z}) \to C^o_*(K;\mathbb{Z})$ that is determined in degree 0 by the formula*

$$\Phi([v]) := (v) \qquad \textit{for all vertices } v \textit{ of } K,$$

and moreover, natural chain maps with this property are unique up to chain homotopy.

LEMMA 32.8. *Natural chain maps $C^o_*(K;\mathbb{Z}) \to C^o_*(K;\mathbb{Z})$ matching the identity map in degree 0 are unique up to chain homotopy.*

LEMMA 32.9. *Natural chain maps $C^\Delta_*(K;\mathbb{Z}) \to C^\Delta_*(K;\mathbb{Z})$ matching the identity map in degree 0 are unique up to chain homotopy.*

Notice that the statements of the last two lemmas only involve uniqueness, not existence; the existence is clear in both cases because the identity map is a chain map that satisfies the required properties. This trivial observation is used in the following proof.

PROOF OF THEOREM 32.6. Writing $\Psi := \Psi_K$, the uniqueness up to chain homotopy in Lemmas 32.8 and 32.9 implies that if we are working with integer coefficients, $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are both chain homotopic to the identity, so that the chain map Φ from Lemma 32.7 is a chain homotopy inverse for Ψ. The validity of this result extends to arbitrary coefficients for relatively trivial algebraic reasons explained in Remark 32.10 below. ☐

REMARK 32.10. Here is why in the proof of Theorem 32.6, it suffices to consider chain complexes with integer coefficients. The three lemmas above provide chain maps between chain complexes with integer coefficients, but the resulting formulas for these maps on the canonical generators of $C^o_*(K;\mathbb{Z})$ and $C^\Delta_*(K;\mathbb{Z})$ determine via linearity chain maps on $C^o_*(K;G)$ and $C^\Delta_*(K;G)$ for any coefficient group G. The same applies to chain homotopies, e.g. if $h : C^o_*(K;\mathbb{Z}) \to C^o_{*+1}(K;\mathbb{Z})$ is a chain homotopy between $\Phi \circ \Psi_K : C^o_*(K;\mathbb{Z}) \to C^o_*(K;\mathbb{Z})$ and the identity, then it determines via linearity a chain homotopy $h : C^o_*(K;G) \to C^o_{*+1}(K;G)$ between $\Phi \circ \Psi_K : C^o_*(K;G) \to C^o_*(K;G)$ and the identity for any choice of coefficients G.

The proofs of Lemmas 32.7, 32.8 and 32.9 are very similar, and are based on an idea known as the method of **acyclic models**. We shall carry out the details only for Lemma 32.7.

PROOF OF LEMMA 32.7. For the entirety of this proof, we assume

$$G := \mathbb{Z}$$

and omit the coefficient group from the notation. We shall prove by induction on the degree $n \geqslant 0$ that it is possible to construct homomorphisms $\Phi_L : C^\Delta_n(L) \to C^o_n(L)$ for every subcomplex $L \subset K$ such that Φ_L is the restriction of Φ_K to $C^\Delta_n(L) \subset C^\Delta_n(K)$ and the chain map relation $\Phi_L\partial = \partial\Phi_L$ is satisfied. It would of course suffice to construct Φ_K such that it sends $C^\Delta_n(L) \to C^o_n(L)$ for every subcomplex $L \subset K$ and then define Φ_L as the restriction, but in practice, we shall do things the other way around, and define Φ_L first for a special class of subcomplexes such that the definition of Φ_K is then uniquely determined.

The beginning of the induction is to define $\Phi_K : C_0^\Delta(L) \to C_0^o(L) : [v] \mapsto (v)$ as specified in the statement of the lemma.

For a given $n \geqslant 1$, we then assume that $\Phi_K : C_k^\Delta(K) \to C_k^o(K)$ has already been defined for every $k \leqslant n-1$ such that it sends $C_k^\Delta(L)$ to $C_k^o(L)$ for every subcomplex $L \subset K$ and satisfies $\Phi_K \partial = \partial \Phi_K$. The idea for the inductive step is now to first define $\Phi_L : C_n^\Delta(L) \to C_n^o(L)$ for a specific class of "model" subcomplexex $L \subset K$, which will determine $\Phi_K : C_n^\Delta(K) \to C_n^o(K)$ via the naturality condition. The model complexes are defined as follows: For any n-simplex $\sigma = \{v_0, \dots, v_n\}$ of K, let $L_\sigma \subset K$ denote the subcomplex that contains only σ and all its faces. Note that since $n \geqslant 1$, L_σ can be identified with the cone of an $(n-1)$-dimensional complex as in Example 32.5, so Lemma 32.4 implies that both versions of the augmented simplicial chain complex for L_σ are acyclic; this is why L_σ is called an "acyclic model". Now, there is only one generator $\sigma = [v_0, \dots, v_n] \in C_n^\Delta(L_\sigma)$, so $\Phi_{L_\sigma} : C_n^\Delta(L_\sigma) \to C_n^o(L_\sigma)$ will be determined as soon as we choose a value for $\tau := \Phi_{L_\sigma}(\sigma) \in C_n^o(L_\sigma)$, which must be required to satisfy

$$\partial\tau = \partial\Phi_{L_\sigma}(\sigma) = \Phi_{L_\sigma}(\partial\sigma) \in C_{n-1}^o(L_\sigma).$$

The right hand side of this expression has already been defined due to the inductive hypothesis. Moreover, it is a cycle in the augmented chain complex $\widetilde{C}_*^o(L_\sigma)$, since

$$\partial\Phi_{L_\sigma}(\partial\sigma) = \Phi_{L_\sigma}(\partial^2\sigma) = 0 \in \widetilde{C}_{n-2}^o(L_\sigma),$$

where we should clarify that in the case $n = 1$, the operator ∂ acting on 0-chains is actually the augmentation $\epsilon_* : C_0^o(L_\sigma; \mathbb{Z}) \to \mathbb{Z}$. Since $\widetilde{C}_*^o(L_\sigma)$ is acyclic, it follows that $\Phi_{L_\sigma}(\partial\sigma)$ is also a boundary, and we can therefore define $\Phi_{L_\sigma}(\sigma)$ to be any choice of element $\tau \in C_n^o(L_\sigma)$ such that $\partial\tau = \Phi_{L_\sigma}(\partial\sigma)$.

Having made such choices and defined $\Phi_{L_\sigma} : C_n^\Delta(L_\sigma) \to C_n^o(L_\sigma)$ for the model subcomplex $L_\sigma \subset K$ corresponding to each n-simplex σ of K, we observe now that there is a unique definition of $\Phi_K : C_n^\Delta(K) \to C_n^o(K)$ that has the correct restriction to all of these subcomplexes, and it automatically satisfies both the chain map relation and the naturality condition.

The construction of Φ_K beyond degree zero involved some arbitrary choices, so it remains to show that any other natural chain map Φ'_K that matches Φ_K in degree zero is chain homotopic to it. We shall use a similar inductive argument to construct homomorphisms $h_K : C_k^\Delta(K) \to C_{k+1}^o(K)$ that satisfy the chain homotopy relation $\partial h_K + h_K \partial = \Phi'_K - \Phi_K$, and here as well it will be convenient to impose a naturality condition, namely that h_K has a well-defined restriction $h_L : C_k^\Delta(L) \to C_{k+1}^o(L)$ for every subcomplex $L \subset K$. To start the induction, it suffices to define $h_K : C_0^\Delta(K) \to C_1^o(K)$ as the trivial homomorphism since $\Phi_K = \Phi'_K$ on $C_0^\Delta(K)$. Now assume that h_K and its restrictions h_L satisfying the chain homotopy relation have already been defined on chains of degree $k \leqslant n-1$ for some $n \geqslant 1$. For each n-simplex $\sigma = \{v_0, \dots, v_n\}$ of K, we again consider the corresponding model subcomplex $L_\sigma \subset K$, and define $h_{L_\sigma} : C_n^\Delta(L_\sigma) \to C_{n+1}^o(L_\sigma)$ so that it sends the unique generator $\sigma = [v_0, \dots, v_n] \in C_n^\Delta(L_\sigma)$ to some element $\tau := h_{L_\sigma}(\sigma) \in C_{n+1}^o(L_\sigma)$ satisfying

$$\partial\tau = \partial h_{L_\sigma}(\sigma) = -h_{L_\sigma}(\partial\sigma) + \Phi'_{L_\sigma}(\sigma) - \Phi_{L_\sigma}(\sigma) \in C_n^o(L_\Sigma).$$

This is possible due to acyclicity, since the inductive hypothesis implies that the right hand side is a cycle:

$$\begin{aligned}\partial\left(-h_{L_\sigma}(\partial\sigma) + \Phi'_{L_\sigma}(\sigma) - \Phi_{L_\sigma}(\sigma)\right) &= \left(-\partial h_{L_\sigma} + \Phi'_{L_\sigma} - \Phi_{L_\sigma}\right)(\partial\sigma) \\ &= h_{L_\sigma}(\partial\partial\sigma) = 0.\end{aligned}$$

Having extended h_{L_σ} to degree n for each of the model subcomplexes $L_\sigma \subset K$, there is again a unique definition of $h_K : C_n^\Delta(K) \to C_{n+1}^o(K)$ that has the correct restriction to each of these subcomplexes, and it automatically satisfies the chain homotopy relation. □

The proofs of Lemmas 32.8 and 32.9 are similar, but shorter since one only needs to construct chain homotopies, the existence of suitable chain maps being obvious. Lemma 32.8 also requires the knowledge that $\widetilde{C}_*^o(L_\sigma)$ is an acyclic chain complex for each of the model complexes $L_\sigma \subset K$, while for Lemma 32.9, one must instead use the fact that $\widetilde{C}_*^\Delta(L_\sigma)$ is acyclic.

32.4. Relative homology. In §28.3 we saw that there is a relative version of bordism theory defined for pairs of spaces $(X, A) \in \mathsf{Top}^{\mathrm{rel}}$, with long exact sequences that relate the relative bordism groups of (X, A) to the absolute bordism groups of X and A. Something similar is true in all versions of homology theory; let's discuss briefly how it works in simplicial homology.

A **simplicial pair** (K, L) is a simplicial complex K together with a subcomplex $L \subset K$, and a **map of simplicial pairs** $f : (K, L) \to (K', L')$ is a simplicial map $f : K \to K'$ that sends L into L' and thus also defines a simplicial map $L \to L'$. Let us denote by $\mathsf{Simp}^{\mathrm{rel}}$ the category whose objects are simplicial pairs and whose morphisms are maps of simplicial pairs. We can identify the category Simp of simplicial complexes with the subcategory

$$\mathsf{Simp} \subset \mathsf{Simp}^{\mathrm{rel}}$$

consisting of pairs of the form $(K, \varnothing)$. The following definition makes sense because for any subcomplex $L \subset K$, the generators of $C_*^\bullet(L)$ are also generators of $C_*^\bullet(K)$, thus making the chain complex $C_*^\bullet(L)$ into a subcomplex of $C_*^\bullet(K)$.

Definition 32.11. The (ordered or oriented) **relative simplicial homology** of a simplicial pair (K, L) with coefficients in G is defined in each degree $n \in \mathbb{Z}$ as the homology of the quotient chain complex $C_*^\bullet(K; G)/C_*^\bullet(L; G)$, thus

$$H_n^\bullet(K, L) = H_n^\bullet(K, L; G) := H_n\big(C_*^\bullet(K, L; G)\big), \qquad \text{where}$$
$$C_*^\bullet(K, L) = C_*^\bullet(K, L; G) := C_*^\bullet(K; G)\Big/C_*^\bullet(L; G).$$

Relative simplicial homology defines functors

$$H_n^\bullet : \mathsf{Simp}^{\mathrm{rel}} \to R\text{-}\mathsf{Mod}$$

in a straightforward way: any map of simplicial pairs $f : (K, L) \to (K', L')$ induces a chain map $f_* : C_*^\bullet(K) \to C_*^\bullet(K')$ that also sends $C_*^\bullet(L)$ to $C_*^\bullet(L')$ and thus descends to the quotients as a chain map $f_* : C_*^\bullet(K, L) \to C_*^\bullet(K, L)$, inducing maps

$$H_n^\bullet(K, L) \xrightarrow{f_*} H_n^\bullet(K', L')$$

for each n. In keeping with the identification of Simp with a subcomplex of $\mathsf{Simp}^{\mathrm{rel}}$, we observe that $H_n(K, \varnothing)$ is the same thing as $H_n(K)$.

Elements $[c] \in H_n^\bullet(K, L)$ can be represented by **relative n-cycles**

$$c \in C_n^\bullet(K) \qquad \text{such that} \qquad \partial c \in C_{n-1}^\bullet(L).$$

Here, the condition $\partial c \in C_{n-1}^\bullet(L)$ means that the image of c under the quotient projection $C_n^\bullet(K) \to C_n^\bullet(K, L)$ is a cycle, and we understand $[c] \in H_n^\bullet(K, L)$ to mean the homology class represented by that cycle. Two relative n-cycles $a, b \in C_n^\bullet(K)$ then represent the same relative homology class in $H_n^\bullet(K, L)$ if and only if $a - b = \partial c + d$ for some $c \in C_{n+1}^\bullet(K)$ and $d \in C_n^\bullet(L)$. There is a natural homomorphism defined for each $n \geqslant 1$ by

$$H_n^\bullet(K, L) \xrightarrow{\partial_*} H_{n-1}^\bullet(L) : [c] \mapsto [\partial c].$$

Note that, in spite of appearances, the class $[\partial c] \in H^{\bullet}_{n-1}(L)$ in this expression need not be trivial, because c is an n-chain in K, but might not be an n-chain in L.

THEOREM 32.12. *Given a simplicial pair (K, L), let $i : L \hookrightarrow K$ and $j : (K, \varnothing) \hookrightarrow (K, L)$ denote the obvious inclusion maps. Then the sequence*

$$\ldots \longrightarrow H^{\bullet}_{n+1}(K, L) \xrightarrow{\partial_*} H^{\bullet}_n(L) \xrightarrow{i_*} H^{\bullet}_n(K) \xrightarrow{j_*} H^{\bullet}_n(K, L) \xrightarrow{\partial_*} H^{\bullet}_{n-1}(L) \xrightarrow{i_*} H^{\bullet}_{n-1}(K) \longrightarrow \ldots$$
$$\longrightarrow H^{\bullet}_0(L) \xrightarrow{i_*} H^{\bullet}_0(K) \xrightarrow{j_*} H^{\bullet}_0(K, L) \longrightarrow 0$$

is exact.

It is not hard to verify the exactness of the sequence in this theorem explicitly, but there is also an underlying algebraic phenomenon that deserves more attention. Since $C^{\bullet}_*(K, L)$ is a quotient, every simplicial pair (K, L) gives rise to an obvious short exact sequence

$$0 \to C^{\bullet}_*(L) \overset{i_*}{\hookrightarrow} C^{\bullet}_*(K) \overset{j_*}{\to} C^{\bullet}_*(K, L) \to 0,$$

in which each term is a chain complex and the maps between them are chain maps. The inclusion $C^{\bullet}_*(L) \hookrightarrow C^{\bullet}_*(K)$ of chain complexes is in fact the chain map i_* induced by the inclusion $i : L \hookrightarrow K$ of simplicial complexes, and since $j : (K, \varnothing) \hookrightarrow (K, L)$ is actually the identity map, the quotient projection $C^{\bullet}_*(K) \to C^{\bullet}_*(K)/C^{\bullet}_*(L)$ can similarly be understood as the chain map $j_* : C^{\bullet}_*(K) \to C^{\bullet}_*(K, L)$ induced by j. Algebraically, it turns out that short exact sequences of chain complexes and chain maps *always* give rise to long exact sequences relating their homology groups:

PROPOSITION 32.13. *Suppose $0 \to A_* \overset{f}{\to} B_* \overset{g}{\to} C_* \to 0$ is a short exact sequence of chain complexes and chain maps. Then for each $n \in \mathbb{Z}$ there exists a so-called* ***connecting homomorphism*** *$\partial_* : H_n(C_*) \to H_{n-1}(A_*)$ such that the sequence*

$$\ldots \xrightarrow{\partial_*} H_{n+1}(A_*) \xrightarrow{f_*} H_{n+1}(B_*) \xrightarrow{g_*} H_{n+1}(C_*)$$
$$\xrightarrow{\partial_*} H_n(A_*) \xrightarrow{f_*} H_n(B_*) \xrightarrow{g_*} H_n(C_*)$$
$$\xrightarrow{\partial_*} H_{n-1}(A_*) \xrightarrow{f_*} H_{n-1}(B_*) \xrightarrow{g_*} H_{n-1}(C_*) \xrightarrow{\partial_*} \ldots$$

is exact. Moreover, this result is functorial in the following sense: suppose we are given another triple of chain complexes A'_*, B'_* and C'_*, with a commuting diagram

$$\begin{tikzcd}
0 \arrow[r] & A_* \arrow[r, "f"] \arrow[d, "\alpha"] & B_* \arrow[r, "g"] \arrow[d, "\beta"] & C_* \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & A'_* \arrow[r, "f'"] & B'_* \arrow[r, "g'"] & C'_* \arrow[r] & 0
\end{tikzcd}$$

in which all maps are chain maps and the bottom row is also exact, and we denote the resulting connecting homomorphisms by $\partial'_* : H_n(C'_*) \to H_{n-1}(A'_*)$. Then the diagram

$$\begin{tikzcd}
\ldots \arrow[r] & H_{n+1}(C_*) \arrow[r, "\partial_*"] \arrow[d, "\gamma_*"] & H_n(A_*) \arrow[r, "f_*"] \arrow[d, "\alpha_*"] & H_n(B_*) \arrow[r, "g_*"] \arrow[d, "\beta_*"] & H_n(C_*) \arrow[r, "\partial_*"] \arrow[d, "\gamma_*"] & H_{n-1}(A_*) \arrow[r] \arrow[d, "\alpha_*"] & \ldots \\
\ldots \arrow[r] & H_{n+1}(C'_*) \arrow[r, "\partial'_*"] & H_n(A'_*) \arrow[r, "f'_*"] & H_n(B'_*) \arrow[r, "g'_*"] & H_n(C'_*) \arrow[r, "\partial'_*"] & H_{n-1}(A'_*) \arrow[r] & \ldots
\end{tikzcd}$$

also commutes.

The proof of this result is by "diagram chasing," which we already saw examples of in Proposition 28.22 and Exercise 28.8 (the five-lemma). Let's do the first step, which is to write down a reasonable candidate for the map $\partial_* : H_n(C_*) \to H_{n-1}(A_*)$. We are given a commuting diagram of the form

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \\
0 \arrow[r] & A_n \arrow[r, "f"] \arrow[d, "\partial"] & B_n \arrow[r, "g"] \arrow[d, "\partial"] & C_n \arrow[r] \arrow[d, "\partial"] & 0 \\
0 \arrow[r] & A_{n-1} \arrow[r, "f"] \arrow[d, "\partial"] & B_{n-1} \arrow[r, "g"] \arrow[d, "\partial"] & C_{n-1} \arrow[r] \arrow[d, "\partial"] & 0 \\
0 \arrow[r] & A_{n-2} \arrow[r, "f"] \arrow[d, "\partial"] & B_{n-2} \arrow[r, "g"] \arrow[d, "\partial"] & C_{n-2} \arrow[r] \arrow[d, "\partial"] & 0 \\
 & \vdots & \vdots & \vdots &
\end{tikzcd}$$

in which every column is a chain complex and every row is exact. Given $[c] \in H_n(C_*)$, choose a representative $c \in C_n$, which necessarily satisfies $\partial c = 0$. We would like to find some element $a \in A_{n-1}$ that satisfies $\partial a = 0$ so that we can set $\partial_*[c] := [a]$. The idea is to use whatever information the diagram gives us to forge a path from C_n to A_{n-1}. To start with, the exactness of the top row implies that g is surjective, so choose $b \in B_n$ with $g(b) = c$. Since $\partial c = 0$ and the diagram commutes, we also know $\partial g(b) = g(\partial b) = 0$, and exactness of the middle row then implies $\partial b = f(a)$ for some $a \in A_{n-1}$. To see that a is a cycle, we use commutativity again and observe $f(\partial a) = \partial f(a) = \partial\partial b = 0$, and since the bottom row is exact, f is injective, so this implies $\partial a = 0$. We can therefore sensibly set $\partial_*[c] = [a]$, and step 1 of the proof is complete.

There are still several things to check: steps 2 through 4000 consist of first verifying that the definition of $\partial_* : H_n(C_*) \to H_{n-1}(A_*)$ we just proposed does not depend on any of the choices we made (e.g. of the representative $c \in C_n$ and the element $b \in g^{-1}(c)$), and after that, we still need to show that the sequence of homology groups really is exact. All of this follows by the same style of diagram chasing—it becomes a bit tedious at some point, but it is not fundamentally difficult. If you haven't done it before, I recommend finding a quiet evening to do so once, so that you never have to do it again.

Similarly, it is not hard to see why the "functoriality" aspect of the statement is true once you have understood the basic idea of diagram chasing. Functoriality in this situation amounts to the statement that there exist natural definitions of categories whose objects are short exact sequences of chain complexes or long exact sequences of R-modules, with morphisms defined in each case via commutative diagrams, such that Proposition 32.13 produces a functor from the former category to the latter. See Exercise 32.3 for a precise formulation in these terms. Exercise 32.2 shows moreover that applying Proposition 32.13 to the short exact sequence $0 \to C_*^\bullet(L) \to C_*^\bullet(K) \to C_*^\bullet(K, L) \to 0$ for a simplicial pair (K, L) produces the same connecting homomorphism as in the statement of Theorem 32.12.

32.5. Exercises.

Exercise 32.1. Carry out the rest of the details of the diagram chase to prove the exactness of the sequence in Proposition 32.13.

Exercise 32.2. Show that for any simplicial pair (K, L), the connecting homomorphisms $\partial_* : H_n^\bullet(K, L) \to H_{n-1}^\bullet(L)$ that arise by plugging the short exact sequence of simplicial chain complexes $0 \to C^\bullet_*(L) \hookrightarrow C^\bullet_*(K) \to C^\bullet_*(K, L) \to 0$ into Proposition 32.13 are given by the formula $\partial_*[c] = \partial c$ for any relative n-cycle $c \in C^\bullet_n(K)$ in (K, L).

Exercise 32.3. Consider the categories Short and Long, defined as follows. Objects in Short are short exact sequences of chain complexes $0 \to A_* \xrightarrow{f} B_* \xrightarrow{g} C_* \to 0$ of R-modules, with a morphism from this object to another object $0 \to A'_* \xrightarrow{f'} B'_* \xrightarrow{g'} C'_* \to 0$ defined as a triple of chain maps $A_* \xrightarrow{\alpha} A'_*$, $B_* \xrightarrow{\beta} B'_*$ and $C_* \xrightarrow{\gamma} C'_*$ such that the following diagram commutes:

$$\begin{tikzcd}
0 \arrow[r] & A_* \arrow[r, "f"] \arrow[d, "\alpha"] & B_* \arrow[r, "g"] \arrow[d, "\beta"] & C_* \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & A'_* \arrow[r, "f'"] & B'_* \arrow[r, "g'"] & C_* \arrow[r] & 0
\end{tikzcd}$$

The objects in Long are long exact sequences of $\mathbb{Z}$-graded R-modules $\ldots \to C_{n+1} \xrightarrow{\delta} A_n \xrightarrow{F} B_n \xrightarrow{G} C_n \xrightarrow{\delta} A_{n-1} \to \ldots$, with morphisms from this to another object $\ldots \to C'_{n+1} \xrightarrow{\delta'} A'_n \xrightarrow{F'} B'_n \xrightarrow{G'} C'_n \xrightarrow{\delta'} A'_{n-1} \to \ldots$ defined as triples of homomorphisms $A_* \xrightarrow{\alpha} A'_*$, $B_* \xrightarrow{\beta} B'_*$ and $C_* \xrightarrow{\gamma} C'_*$ that preserve the $\mathbb{Z}$-gradings and make the following diagram commute:

$$\begin{tikzcd}
\ldots \arrow[r] & C_{n+1} \arrow[r, "\delta"] \arrow[d, "\gamma"] & A_n \arrow[r, "F"] \arrow[d, "\alpha"] & B_n \arrow[r, "G"] \arrow[d, "\beta"] & C_n \arrow[r, "\delta"] \arrow[d, "\gamma"] & A_{n-1} \arrow[r] \arrow[d, "\alpha"] & \ldots \\
\ldots \arrow[r] & C'_{n+1} \arrow[r, "\delta'"] & A'_n \arrow[r, "F'"] & B'_n \arrow[r, "G'"] & C'_n \arrow[r, "\delta'"] & A'_{n-1} \arrow[r] & \ldots
\end{tikzcd}$$

(a) Show that there is a covariant functor $\mathsf{Simp}^{\text{rel}} \to$ Short assigning to each simplicial pair (K, L) its short exact sequence of (ordered or oriented) simplicial chain complexes.

(b) Show that there is also a covariant functor Short → Long assigning to each short exact sequence of chain complexes the corresponding long exact sequence of their homology groups. (Note that this can be composed with the functor in part (a) to define a functor $\mathsf{Simp}^{\text{rel}} \to$ Long.)

33. Singular homology

33.1. Definitions. The immediate disadvantage of simplicial homology is that its definition requires strictly more data than just a topological space: we need to have a triangulation of that space, and it takes considerable effort to see why different triangulations of the same space produce isomorphic homologies. The definition of singular homology resembles that of simplicial homology, but it explicitly removes the need for a triangulation. The price to be paid for this is that the resulting chain complex seems absurdly large: so large, in fact, that one might find it surprising at first that it is ever possible to explicitly compute the singular homology of a space. I advise you not to think too much about this when you first read the definition, as we will subsequently discuss some properties that make computations of singular homology quite a reasonable task.

Definition 33.1. A **singular n-simplex** (*singulärer n-Simplex*) in a topological space X is defined to be a continuous map $\sigma : \Delta^n \to X$. Let

$$\mathcal{K}_n(X) := \left\{ \sigma : \Delta^n \to X \;\middle|\; \sigma \text{ is continuous} \right\}$$

denote the set of all singular n-simplices in X. The **singular chain complex** (*singulärer Kettenkomplex*) $C_*(X) = C_*(X; G)$ of X with coefficients in the R-module G is defined such that

$C_n(X) = 0$ for all $n < 0$, and for $n \geqslant 0$,

$$C_n(X) := \bigoplus_{\sigma \in \mathcal{K}_n(X)} G.$$

The boundary map $\partial : C_n(X) \to C_{n-1}(X)$ for $n \geqslant 1$ is uniquely determined by linearity and the formula

$$\partial\sigma = \sum_{k=0}^{n} (-1)^k \left(\sigma|_{\partial_{(k)}\Delta^n}\right),$$

where the identification (29.1) is used in order to view each term in the summation as a singular $(n-1)$-simplex $\sigma|_{\partial_{(k)}\Delta^n} : \Delta^{n-1} \to X$, making the linear combination an element of $C_{n-1}(X;\mathbb{Z})$. The homology groups of this chain complex form the **singular homology** of X with coefficients in G,

$$H_n(X) = H_n(X;G) := H_n\big(C_*(X;G)\big).$$

There is a fairly obvious way to make

$$C_* : \mathsf{Top} \to \mathsf{Ch}(R\text{-}\mathsf{Mod})$$

into a functor: any continuous map $f : X \to Y$ between spaces induces a unique chain map

$$f_* : C_*(X) \to C_*(Y)$$

determined by linearity and the formula

$$f_*(\sigma) := f \circ \sigma$$

for singular simplices $\sigma : \Delta^n \to X$. Composing this functor with $H_n : \mathsf{Ch}(R\text{-}\mathsf{Mod}) \to R\text{-}\mathsf{Mod}$ makes singular homology itself into a collection of functors

$$H_n : \mathsf{Top} \to R\text{-}\mathsf{Mod},$$

meaning in particular that continuous maps $f : X \to Y$ induce homomorphisms $f_* : H_n(X) \to H_n(Y)$ for every $n \geqslant 0$.

For a pair of spaces $(X, A) \in \mathsf{Top}^{\text{rel}}$, there is a similarly straightforward extension of the definitions above to the notion of **relative singular homology**

$$\begin{aligned} H_n(X,A) = H_n(X,A;G) &:= H_n\big(C_*(X,A;G)\big), \qquad \text{where} \\ C_*(X,A) = C_*(X,A;G) &:= C_*(X;G)\big/C_*(A;G), \end{aligned}$$

which makes sense because singular simplices in A are also singular simplicies in X, making $C_*(A)$ naturally a subcomplex of $C_*(X)$. We have $H_n(X, \varnothing) = H_n(X)$ for all spaces X, and relative singular homology can thus be regarded as an extension of the functor $H_n : \mathsf{Top} \to R\text{-}\mathsf{Mod}$ over the larger category $\mathsf{Top}^{\text{rel}}$, in which maps of pairs $f : (X, A) \to (Y, B)$ induce chain maps $f_* : C_*(X) \to C_*(Y)$ that descend to the quotients as chain maps $f_* : C_*(X, A) \to C_*(Y, B)$ and thus induce homomorphisms $f_* : H_n(X, A) \to H_n(Y, B)$ for all n. As with relative simplicial homology, we can represent relative singular homology classes $[c] \in H_n(X, A)$ via **relative cycles** $c \in C_n(X)$, which are assumed to satisfy $\partial c \in C_{n-1}(A)$, and writing them in this way gives rise to an obvious connecting homomorphism

$$H_n(X, A) \xrightarrow{\partial_*} H_{n-1}(A) : [c] \mapsto [\partial c].$$

33.2. Non-axiomatic properties. Before we get to the purely "formal" (i.e. axiomatic) properties of singular homology, let us discuss a few features it has that other axiomatic homology theories do not.

THEOREM 33.2. *For any space X and any coefficient group G, there is a canonical isomorphism*

$$H_0(X;G) = \bigoplus_{\pi_0(X)} G,$$

where $\pi_0(X)$ is an abbreviation for the set of path-components of X.

The isomorphism in this theorem arises from a pair of convenient coincidences: first, since the standard 0-simplex Δ^0 contains only one point, there is a natural bijection between the set $\mathcal{K}_0(X)$ of singular 0-simplices in X and the set X itself, allowing us to write singular 0-chains as finite linear combinations

$$\sum_i a_i x_i \in C_0(X)$$

of generators $x_i \in X$ with coefficients $a_i \in G$. The second coincidence is that the unit interval $I = [0,1]$, which we normally use for parametrizing paths in X, is homeomorphic to the standard 1-simplex $\Delta^1 \subset I^2$, e.g. via the map

$$I \xrightarrow{\cong} \Delta^1 : t \mapsto (1-t, t). \tag{33.1}$$

This is of course not the only possible choice of such a homeomorphism, but we will use it consistently in this course, for the following reason. The map (33.1) matches boundary points via the correspondence

$$\partial I \ni 0 \mapsto \partial_{(1)}\Delta^1 \subset \partial\Delta^1, \qquad \partial I \ni 1 \mapsto \partial_{(0)}\Delta^1 \subset \partial\Delta^1,$$

which may seem backwards when you see it for the first time, but if you recall the way in which signs were associated to the various boundary faces of Δ^n in our definition of the boundary operator $\partial : C_n(X) \to C_{n-1}(X)$, you might recognize that this particular correspondence is consistent with certain orientation conventions in differential geometry, where the standard orientation of the 1-manifold $I \subset \mathbb{R}$ induces a positive boundary orientation on $1 \in \partial I$ and a negative boundary orientation on $0 \in \partial I$. This detail is unimportant for our present purposes, but what matters is that if we use (33.1) to identify singular 1-simplices in X with paths $\gamma : I \to X$ and likewise identify singular 0-simplices with points $x \in X$ in the canonical way, then the operator $\partial : C_1(X) \to C_0(X)$ is now determined by the formula

$$\partial\gamma = \gamma(1) - \gamma(0). \tag{33.2}$$

This tells you why two 0-cycles of the form $mx, my \in C_0(X)$ for $m \in G$ and $x, y \in X$ will always be homologous if x and y lie in the same path-component, and from there it is not a difficult exercise to find an explicit isomorphism $H_0(X) \cong \bigoplus_{\pi_0(X)} G$.

For any choice of base point $p \in X$, the identification (33.1) between I and Δ^1 also gives rise to a natural homomorphism

$$h : \pi_1(X,p) \to H_1(X;\mathbb{Z}) \tag{33.3}$$

sending the homotopy class of the loop $\gamma : I \to X$ to the homology class that it represents when regarded as a singular 1-chain with integer coefficients; note that by (33.2), this 1-chain is a cycle because $\gamma : I \to X$ has the same start and end point. The map (33.3) is called the **Hurewicz homomorphism**, and the proof that it is well defined (see e.g. Exercise 22.12 from last semester's *Topologie I* course) relies on several straightforward lemmas, showing for instance that any two homotopic loops based at p give rise to homologous 1-cycles, and the 1-cycle arising from a concatenation of two loops is homologous to the sum of the two corresponding 1-cycles.

Since $H_1(X;\mathbb{Z})$ is abelian, the Hurewicz map automatically vanishes on the commutator subgroup of $\pi_1(X,p)$, so it descends to a map of the abelianization of $\pi_1(X,p)$ to $H_1(X;\mathbb{Z})$.

THEOREM 33.3. *If X is path-connected, then the Hurewicz map (33.3) descends to the abelianization of $\pi_1(X) := \pi_1(X,p)$ as an isomorphism*

$$\pi_1(X)\Big/[\pi_1(X),\pi_1(X)] \xrightarrow{\cong} H_1(X;\mathbb{Z}).$$

One can prove Theorem 33.3 by writing down an inverse map that transforms any singular 1-cycle (viewed as a formal sum of paths whose end points must satisfy some matching conditions in order to produce a cycle) into a loop based at p by concatenating the associated paths. There are typically many ways that this can be done, but the ambiguity turns out to lie in the commutator subgroup $[\pi_1(X,p),\pi_1(X,p)]$; see last semester's Exercise 22.12 for further hints.

The third property I want to mention is a relationship between simplicial and singular homology. Suppose $K = (V,S)$ is a simplicial complex, with polyhedron $|K|$. There is then a natural chain map

$$C_*^o(K) \to C_*(|K|)$$

defined by associating to each generator $(v_0,\ldots,v_n)$ in degree n of the ordered simplicial chain complex the unique singular n-simplex $\sigma : \Delta^n \to |K|$ that extends to a linear map $\mathbb{R}^{n+1} \to \mathbb{R}^V$ sending the standard basis of $\mathbb{R}^{n+1}$ to the vectors $e_{v_0},\ldots,e_{v_n} \in \mathbb{R}^V$. As usual, the word "natural" has a precise meaning here, and the map $C_*^o(K) \to C_*(|K|)$ can be described as a natural transformation between two functors $\mathsf{Simp} \to \mathsf{Ch}(R\text{-}\mathsf{Mod})$. Letting chain maps descend to maps between homology groups, we obtain natural homomorphisms

$$H_n^o(K) \to H_n(|K|).$$

In light of the natural isomorphisms $H_n^o(K) \xrightarrow{\cong} H_n^\Delta(K)$, we also obtain from this natural homomorphisms $H_n^\Delta(K) \to H_n(|K|)$, and we will see when we study cellular homology that the latter is also an isomorphism.

One useful application of this relationship is a construction of fundamental cycles in singular homology: If $M \cong |K|$ is a compact triangulated n-manifold, with a choice of admissible ordering for the underlying simplicial complex, then feeding the resulting fundamental cycle $c_M \in C_n^o(K;\mathbb{Z}_2)$ into the natural chain map $C_n^o(K;\mathbb{Z}_2) \to C_n(M;\mathbb{Z}_2)$ produces a singular fundamental cycle

$$c_M \in C_n(M;\mathbb{Z}_2) \qquad \text{such that} \qquad \partial c_M = c_{\partial M} \in C_{n-1}(\partial M;\mathbb{Z}_2) \subset C_{n-1}(M;\mathbb{Z}_2).$$

If the triangulation is also oriented, then this can all also be done with integer coefficients, producing an integral fundamental cycle

$$c_M \in C_n(M;\mathbb{Z}) \qquad \text{such that} \qquad \partial c_M = c_{\partial M} \in C_{n-1}(\partial M;\mathbb{Z}) \subset C_{n-1}(M;\mathbb{Z}).$$

33.3. The axioms.

Here is the main result of this lecture.

THEOREM 33.4. *For any R-module G, the functors $H_n(\cdot;G) : \mathsf{Top}^{\text{rel}} \to R\text{-}\mathsf{Mod}$ and connecting homomorphisms $\partial_* : H_n(X,A;G) \to H_{n-1}(A;G)$ defined for all $(X,A) \in \mathsf{Top}^{\text{rel}}$ and $n \in \mathbb{Z}$ satisfy the axioms of a homology theory (in the sense of Eilenberg-Steenrod) with coefficient group G.*

Let's first dispense with the axioms that are easy exercises. Since a one-point space $\{*\}$ admits only one singular n-simplex $\sigma : \Delta^n \to \{*\}$ for each $n \geqslant 0$, a computation completely analogous to Exercise 29.4 shows that

$$H_n(\{*\};G) \cong \begin{cases} 0 & \text{if } n \neq 0, \\ G & \text{if } n = 0. \end{cases}$$

This shows that $H_* := H_*(\cdot; G)$ satisfies the *dimension* axiom, and moreover, the isomorphism $H_0(\{*\}; G) \cong G$ is canonical. The *additivity* axiom is a similarly straightforward consequence of the definitions.

Now for the interesting part.

PROPOSITION 33.5 (the homotopy axiom). *For any two homotopic maps of pairs $f, g : (X, A) \to (Y, B)$, the induced homomorphisms $f_*, g_* : H_n(X, A) \to H_n(Y, B)$ on singular homology are identical.*

PROOF. We consider first the case of absolute homology, so assume $h : I \times X \to Y$ is a homotopy between $f := h(0, \cdot)$ and $g := h(1, \cdot)$. For each $n \geqslant 0$, there is a unique homomorphism $h_\# : C_n(X) \to C_{n+1}(Y)$ determined by linearity and the following formula for $h_\#(\sigma) \in C_{n+1}(Y; \mathbb{Z})$ on an arbitrary singular n-simplex $\sigma : \Delta^n \to X$: we use the maps

$$I \times \Delta^n \xrightarrow{\operatorname{Id} \times \sigma} I \times X \xrightarrow{h} Y,$$

together with the integral fundamental cycle $c_{I\times\Delta^n} \in C_{n+1}(I \times \Delta^n; \mathbb{Z})$ arising from the oriented triangulation of $I \times \Delta^n \cong \Delta^1 \times \Delta^n$ described in §31.2, to define

$$h_\#(\sigma) := h_*(\operatorname{Id} \times \sigma)_* c_{I\times\Delta^n} \in C_{n+1}(Y; \mathbb{Z}).$$

One now deduces from the formula for $\partial c_{I\times\Delta^n}$ that $h_\# : C_*(X) \to C_{*+1}(Y)$ is a chain homotopy between f_* and g_*.

Extending this result to the setting of a homotopy $h : (I \times X, I \times A) \to (Y, B)$ between two maps of pairs $f, g : (X, A) \to (Y, B)$ requires only the extra observation that since $h(I \times A) \subset B$, the chain homotopy $h_\#$ constructed above descends to the quotient as a chain homotopy $C_*(X, A) \to C_{*+1}(Y, B)$ between the two chain maps $f_*, g_* : C_*(X, A) \to C_*(Y, B)$. □

PROPOSITION 33.6 (the exactness axiom). *For any pair of spaces $(X, A) \in \mathsf{Top}^{\text{rel}}$ with inclusion maps $i : A \hookrightarrow X$ and $j : (X, \varnothing) \hookrightarrow (X, A)$, the sequence*

$$\ldots \longrightarrow H_n(A) \xrightarrow{i_*} H_n(X) \xrightarrow{j_*} H_n(X, A) \xrightarrow{\partial_*} H_{n-1}(A) \longrightarrow \ldots \longrightarrow H_0(X, A) \longrightarrow 0$$

is exact.

PROOF. This is a straightforward consequence of Proposition 32.13 and the obvious short exact sequence of chain complexes

$$0 \longrightarrow C_*(A) \xrightarrow{i_*} C_*(X) \xrightarrow{j_*} C_*(X, A) \longrightarrow 0,$$

one only needs to check that the connecting homomorphism produced by the diagram chase in the proof of Proposition 32.13 is the specific map $H_n(X, A) \to H_{n-1}(A) : [c] \mapsto [\partial c]$. □

Recall that for the *excision* axiom formulated in Lecture 28, the hypothesis was that $B \subset A \subset X$ and there exists a continuous function $u : X \to I$ that "separates" B from $X \backslash A$ in the sense that $u|_B \equiv 0$ and $u|_{X\backslash A} \equiv 1$. In singular homology, it suffices to work with a slightly weaker variant of this hypothesis.

PROPOSITION 33.7 (the excision axiom). *Assume $B \subset A \subset X$ such that the closure of B is contained in the interior of A. Then the inclusion of pairs $i : (X\backslash B, A\backslash B) \hookrightarrow (X, A)$ induces an isomorphism $i_* : H_n(X\backslash B, A\backslash B) \xrightarrow{\cong} H_n(X, A)$ for every n.*

The proof requires a bit of preparation. For each $n \geqslant 0$, there is a unique homomorphism

$$S : C_n(X) \to C_n(X)$$

that is determined by linearity and the formula

$$S(\sigma) := \sigma_* c_{\Delta^n} \in C_n(X;\mathbb{Z}),$$

where $\sigma : \Delta^n \to X$ is an arbitrary singular n-simplex and $c_{\Delta^n} \in C_n(\Delta^n;\mathbb{Z})$ is the integral fundamental cycle defined via barycentric subdivision of Δ^n. We will refer to this as the **subdivision operator** on the singular chain complex.

LEMMA 33.8. *For each $m \in \mathbb{N}$, the mth iterate $S^m : C_*(X) \to C_*(X)$ of the subdivision operator is a chain map, and there exists a chain homotopy $h_m : C_*(X) \to C_{*+1}(X)$ between S^m and the identity map. Moreover, for any subspace $A \subset X$, both S^m and h_m preserve the subcomplex $C_*(A) \subset C_*(X)$.*

PROOF. We prove the statement first for $m = 1$. For a given singular n-simplex $\sigma : \Delta^n \to X$, the chain map relation $\partial S(\sigma) = \sigma_* \partial c_{\Delta^n} = \sigma_* c_{\partial\Delta^n} = S(\partial\sigma)$ follows from the inductive nature of the barycentric subdivision algorithm described in §30.4. To see why S is then chain homotopic to the identity, one uses the oriented triangulation of $I \times \Delta^n$ described in §31.4 and its integral fundamental cycle $c_{I\times\Delta^n} \in C_{n+1}(I \times \Delta^n;\mathbb{Z})$ to define, for each $n \geqslant 0$, a homomorphism $h_1 : C_n(X) \to C_{n+1}(X)$ that is given on each generator $\sigma : \Delta^n \to X$ by

$$h_1(\sigma) = (\mathrm{pr}_2)_*(\mathrm{Id} \times \sigma)_* c_{I\times\Delta^n} \in C_{n+1}(X;\mathbb{Z}),$$

with $\mathrm{pr}_2 : I \times X \to X$ denoting the projection to the second factor. Since the triangulation of $I \times \Delta^n$ restricts to $\partial(I \times \Delta^n)$ as the trivial triangulation of $\{0\} \times \Delta^n$, the barycentric subdivision of $\{1\} \times \Delta^n$, and the $(n-1)$-dimensional case of the same triangulation on each face of $I \times \partial\Delta^n$, we have

$$\partial h_1(\sigma) = (\mathrm{pr}_2)_*(\mathrm{Id} \times \sigma)_* \partial c_{I\times\Delta^n} = S(\sigma) - \sigma - h_1(\partial\sigma),$$

and thus the chain homotopy relation $\partial h_1 + h_1 \partial = S - \mathbb{1}$. It is clear from the construction that both S and h_1 preserve $C_*(A) \subset C_*(X)$ for any $A \subset X$.

For arbitrary $m \in \mathbb{N}$, it is now obvious that S^m is also a chain map and is chain homotopic to $\mathbb{1}^m = \mathbb{1}$, but we need to check that there is a chain homotopy h_m that preserves subcomplexes $C_*(A) \subset C_*(X)$. One can see this by writing down an inductive definition of h_m, for which various choices are possible, e.g. $h_m := h_{m-1}S + h_1$ does the job. □

Taking m large enough, the operator S^m can be applied in principle to replace any singular chain $c \in C_n(X)$ with a chain $S^m c \in C_n(X)$ whose constituent singular simplices are as "small" we we like: in particular, if X is covered by the interiors of two subsets

$$X = \mathring{\mathcal{U}} \cup \mathring{\mathcal{V}}, \qquad \mathcal{U}, \mathcal{V} \subset X,$$

then for any given chain $c \in C_n(X)$, taking $m \in \mathbb{N}$ sufficiently large makes the n-chain $S^m c \in C_n(X)$ *decomposable* with respect to this covering, meaning

$$S^m c = u + v \qquad \text{for some} \qquad u \in C_n(\mathcal{U}),\ v \in C_n(\mathcal{V}),$$

because every singular n-simplex in the finite linear combination forming $S^m c$ can be assumed to have its image entirely inside either $\mathcal{U}$ or $\mathcal{V}$. Moreover, if $c \in C_n(X)$ is a cycle, then $S^m c \in C_n(X)$ is also a cycle, and the chain homotopy relation

$$S^m c - c = \partial h_m c + h_m \partial c = \partial h_m c$$

shows that c and $S^m c$ represent the same singular homology class. A relative version of this observation will be used in the proof below, and we can now see the significance of the condition $\bar{B} \subset \mathring{A}$: it means that the interiors of A and $X \backslash B$ form an open covering of X.

PROOF OF PROPOSITION 33.7. Given any class $[c] \in H_n(X, A)$ represented by a relative n-cycle $c \in C_n(X)$, we observe that for each $m \in \mathbb{N}$, the chain $S^m c \in C_n(X)$ satisfies

$$\partial(S^m c) = S^m(\partial c) \in C_{n-1}(A),$$

since the subdivision operator $S : C_*(X) \to C_*(X)$ preserves the subcomplex $C_*(A) \subset C_*(X)$, hence $S^m c$ is also a relative n-cycle. Moreover, the chain homotopy relation $S^m c - c = \partial h_m c + h_m \partial c$ implies $[S^m c] = [c] \in H_n(X, A)$, since $\partial c \in C_{n-1}(A)$ implies $h_m \partial c \in C_n(A)$. With this in mind, since the interiors of A and $X \backslash B$ cover X, we can assume without loss of generality after replacing c by $S^m c$ for some $m \in \mathbb{N}$ sufficiently large that the chain c can be decomposed as

$$c = c_A + c_{X\backslash B} \qquad \text{for some} \qquad c_A \in C_n(A),\ c_{X\backslash B} \in C_n(X\backslash B).$$

Having made this assumption, the fact that $c \in C_n(X, A)$ is a relative n-cycle means $\partial c \in C_{n-1}(A)$ and therefore also $\partial c_{X\backslash B} \in C_n(A)$, so that $c_{X\backslash B}$ is a relative n-cycle in $(X\backslash B, A\backslash B)$, thus representing a class $[c_{X\backslash B}] \in H_n(X\backslash B, A\backslash B)$ that satisfies

$$i_*[c_{X\backslash B}] = [c].$$

This proves that $i_* : H_n(X\backslash B, A\backslash B) \to H_n(X, A)$ is surjective.

To show that $i_* : H_n(X\backslash B, A\backslash B) \to H_n(X, A)$ is injective, suppose $c \in C_n(X\backslash B)$ is a relative n-cycle in $(X\backslash B, A\backslash B)$ representing a class $[c] \in H_n(X\backslash B, A\backslash B)$ with $i_*[c] = 0 \in H_n(X, A)$, which means that if c is viewed as an n-chain in X, we have

$$c = \partial b + a \qquad \text{for some } b \in C_{n+1}(X) \text{ and } a \in C_n(A).$$

By applying S^m to both sides for m sufficiently large, we can assume without loss of generality that b decomposes as

$$b = b_A + b_{X\backslash B} \qquad \text{for some} \qquad b_A \in C_n(A),\ b_{X\backslash B} \in C_n(X\backslash B).$$

We then have $c - \partial b_{X\backslash B} = \partial b_A + a$, in which the left hand side is a chain in $X\backslash B$ and the right hand side is a chain in A, implying that the right hand side is also a chain in $A\backslash B$, and the relation $c = \partial b_{X\backslash B} + (\partial b_A + a)$ therefore implies $[c] = 0 \in H_n(X\backslash B, A\backslash B)$. □

The proof of Theorem 33.4 is now complete, and now that we have established the existence of at least one axiomatic homology theory with any given choice of coefficient group, there are many immediate corollaries, e.g. the Brouwer fixed point theorem (cf. Exercise 28.7). In particular, the computation of $h_*(S^n)$ carried out in Lecture 28 can now be considered a valid computation of singular homology, giving

$$H_k(S^n; G) \cong \begin{cases} G & \text{if } k = 0 \text{ or } k = n, \\ 0 & \text{otherwise.} \end{cases}$$

33.4. Chain-level excision. When we study singular cohomology later in this semester, it will be useful to have a stronger variant of the excision property, one that applies to the singular *chain complex* rather than just to its homology:

THEOREM 33.9. *Assume $B \subset A \subset X$ such that the closure of B is contained in the interior of A. Then the inclusion of pairs $i : (X\backslash B, A\backslash B) \hookrightarrow (X, A)$ induces a chain homotopy equivalence $i_* : C_*(X\backslash B, A\backslash B) \to C_*(X, A)$.*

This obviously implies Proposition 33.7, and we will see in the following that it also more-or-less follows from it, for somewhat nontrivial reasons. The usefulness of Theorem 33.9 will be that it almost immediately implies that similar statements hold after applying certain standard algebraic operations to chain complexes, such as the one that turns homology into cohomology. One gets a hint of this from the following easy observation, which is based on the same trick as Remark 32.10:

Lemma 33.10. *If Theorem 33.9 holds for singular chain complexes with coefficients in $\mathbb{Z}$, then it also holds for arbitrary choices of coefficients.*

Proof. Under the stated hypotheses on $B \subset A \subset X$, assume it is known that the map $i_* : C_*(X \backslash B, A \backslash B; \mathbb{Z}) \to C_*(X, A; \mathbb{Z})$ is a chain homotopy equivalence, which means there exists a chain map $g : C_*(X, A; \mathbb{Z}) \to C_*(X \backslash B, A \backslash B; \mathbb{Z})$, a chain homotopy h_1 between $i_* g : C_*(X, A; \mathbb{Z}) \to C_*(X, A; \mathbb{Z})$ and the identity, and a chain homotopy h_2 between $g i_* : C_*(X \backslash B, A \backslash B; \mathbb{Z}) \to C_*(X \backslash B, A \backslash B; \mathbb{Z})$ and the identity. For any coefficient module G, linearity and the definitions of these maps on the generators of the singular chain complex (i.e. on singular simplices) uniquely determine similar chain maps and chain homotopies that relate $C_*(X \backslash B, A \backslash B; G)$ and $C_*(X, A; G)$ in the same manner. □

With the lemma in mind, our goal is now to prove that Theorem 33.9 holds in the special case $G = \mathbb{Z}$. We will deduce this from some general results about chain complexes in the next subsection.

33.5. Chain contractions and mapping cones. Recall that a chain complex C_* is called *chain contractible* if there exists a chain homotopy of the identity map $C_* \to C_*$ to the trivial chain map $0 : C_* \to C_*$.

Lemma 33.11. *A chain complex of R-modules C_* is chain contractible if and only if there is a splitting of C_n for each $n \in \mathbb{Z}$ into submodules $C_n = A_n \oplus B_n$ such that $C_n \xrightarrow{\partial} C_{n-1}$ vanishes on A_n and maps B_n isomorphically to A_{n-1}.*

Proof. Assume $h : C_* \to C_*$ satisfies $h(C_n) \subset C_{n+1}$ for every $n \in \mathbb{Z}$ and $\partial h + h\partial = \mathbb{1}$. We observe that the homomorphisms ∂h and $h\partial$ in this case are complementary projections, since $\partial^2 = 0$ implies

$$(\partial h)^2 = \partial(h\partial)h = \partial(\mathbb{1} - \partial h)h = \partial h, \qquad \text{and} \qquad (h\partial)^2 = h(\partial h)\partial = h(\mathbb{1} - h\partial)\partial = h\partial.$$

We therefore obtain a splitting $C_* = A_* \oplus B_*$ with $A_* := \operatorname{im}(\partial h)$ and $B_* := \operatorname{im}(h\partial)$, and setting $A_n := A_* \cap C_n$ and $B_n := B_* \cap C_n$ for each $n \in \mathbb{Z}$ gives $C_n = A_n \oplus B_n$. Since $\partial(\partial h) = 0$, ∂ vanishes on A_*; moreover, the definitions of the projections imply that ∂h is the identity map on A_* while $h\partial$ is the identity map on B_*, showing that for each $n \in \mathbb{Z}$, one obtains an inverse of $B_n \xrightarrow{\partial} A_{n-1}$ by composing $A_{n-1} \xrightarrow{h} C_n$ with the projection $C_n \xrightarrow{h\partial} B_n$.

Conversely, if splittings $C_n = A_n \oplus B_n$ with the stated properties are given, then defining $h : C_n \to C_{n+1}$ for each $n \in \mathbb{Z}$ to be trivial on B_n and an inverse of $B_{n+1} \xrightarrow{\partial} A_n$ on A_n gives a chain contraction. □

We can now clarify the advantage of focusing on the case $G = \mathbb{Z}$ in the proof of Theorem 33.9: it is the fact that chain complexes over $\mathbb{Z}$ are *free* abelian groups, thus making the following result applicable.

Lemma 33.12. *A chain complex C_* of free abelian groups is acyclic if and only if it is chain contractible.*

Proof. Assume the chain complex C_* is acyclic and $C_n \subset C_*$ is a free abelian group for each $n \in \mathbb{Z}$. By a basic result in algebra (see e.g. [Lan02, §III.7]), subgroups of free abelian groups are also free, and this applies in particular to the subgroups

$$Z_n := \ker\left(C_n \xrightarrow{\partial_n} C_{n-1}\right).$$

Acyclicity means that the map

$$C_n \xrightarrow{\partial_n} Z_{n-1}$$

is surjective for every n, and we therefore have a short exact sequence

$$0 \to Z_n \hookrightarrow C_n \xrightarrow{\partial} Z_{n-1} \to 0,$$

which splits since Z_{n-1} is free. This splitting identifies C_n with $Z_n \oplus Z_{n-1}$ so that ∂ becomes the projection $Z_n \oplus Z_{n-1} \to Z_{n-1}$, and chain contractibility now follows from Lemma 33.11. □

Remark 33.13. Lemma 33.12 also holds under the hypothesis that each $C_n \subset C_*$ is a free R-module if the underlying ring R is a principal ideal domain: the key detail is that under this assumption, submodules of free R-modules are also free, thus providing the splitting of short exact sequences used in the proof. This fact about principal ideal domains depends in general on Zorn's lemma, so it may seem a bit abstract, but one can also avoid using it if one is willing to assume the chain complex is *bounded* above or below, which also suffices for our present purposes; see Exercise 33.2.

What we need next is a way to deduce that something is a chain homotopy equivalence from the fact that some other complex is chain contractible. The right tool for this is the mapping cone.

A quick digression on simplicial complexes will provide some useful motivation. For a simplicial pair (K, L), the cone CL of L also contains L itself as a subcomplex, so we can define the **cone** of the pair (K, L) as the simplicial complex

$$\mathrm{cone}(K, L) := CL \cup_L K,$$

in which the subcomplexes $L \subset CL$ and $L \subset K$ are identified with each other. The vertices of $\mathrm{cone}(K, L)$ thus consist of the vertices of K plus one extra vertex labelled $*$, while its simplices consist of the simplices of K plus, for each $n \geqslant 0$ and each n-simplex $\{v_0, \ldots, v_n\}$ of L, the $(n+1)$-simplex $\{*, v_0, \ldots, v_n\}$. Topologically, the polyhedron $|\mathrm{cone}(K, L)|$ is a space obtained by attaching $|K|$ to the cone $C|L|$ along $|L|$, thus making the inclusion $|L| \hookrightarrow |\mathrm{cone}(K, L)|$ homotopic to a constant map.

The augmented chain complex $\widetilde{C}^{\Delta}_*(\mathrm{cone}(K, L); \mathbb{Z})$ contains two types of generators. First, since $K \subset \mathrm{cone}(K, L)$ is a subcomplex, there are the generators corresponding to simplices of K, in addition to $1 \in \mathbb{Z} = \widetilde{C}^{\Delta}_{-1}(K; \mathbb{Z})$, making $\widetilde{C}^{\Delta}_*(K; \mathbb{Z})$ a subcomplex of $\widetilde{C}^{\Delta}_*(\mathrm{cone}(K, L); \mathbb{Z})$. Secondly, each oriented simplex $[v_0, \ldots, v_n]$ of L gives rise to a generator $[*, v_0, \ldots, v_n]$ of $\widetilde{C}^{\Delta}_*(\mathrm{cone}(K, L))$, defining for each $n \geqslant -1$ a homomorphism

$$\widetilde{C}^{\Delta}_n(L; \mathbb{Z}) \xrightarrow{j} \widetilde{C}^{\Delta}_{n+1}(\mathrm{cone}(K, L); \mathbb{Z})$$

such that $j[v_0, \ldots, v_n] := [*, v_0, \ldots, v_n]$ and, for the case $n = -1$, $j(1) := [*]$. The map j identifies every n-chain in $\widetilde{C}^{\Delta}_*(L; \mathbb{Z})$ with an $(n+1)$-chain in $\widetilde{C}^{\Delta}_*(\mathrm{cone}(K, L); \mathbb{Z})$, but j is not a chain map and its image is not a subcomplex: instead, we have

$$\partial j[v_0, \ldots, v_n] = \partial[*, v_0, \ldots, v_n] = [v_0, \ldots, v_n] - j(\partial[v_0, \ldots, v_n])$$

and, using $\partial = \epsilon_*$ for the degree 0 part of the augmented chain complex, $\partial j(1) = \epsilon_*[*] = 1 = 1 - j(\partial(1))$. The result is a direct sum decomposition

$$\widetilde{C}^{\Delta}_n(\mathrm{cone}(K, L); \mathbb{Z}) \cong \widetilde{C}^{\Delta}_{n-1}(L; \mathbb{Z}) \oplus \widetilde{C}^{\Delta}_n(K; \mathbb{Z})$$

for each $n \geqslant -1$ such that the boundary map on $\widetilde{C}^{\Delta}_*(\mathrm{cone}(K, L); \mathbb{Z})$ decomposes in block form as

$$\partial = \begin{pmatrix} -\partial^L & 0 \\ i_* & \partial^K \end{pmatrix},$$

where ∂^L and ∂^K denote the boundary maps on the augmented simplicial chain complexes of L and K respectively, and $i : L \hookrightarrow K$ is the inclusion. One can take this as topological motivation for the following algebraic definition.

Definition 33.14. The **mapping cone** of a chain map $f : (A_*, \partial^A) \to (B_*, \partial^B)$ is the chain complex $(\mathrm{cone}(f)_*, \partial)$ with

$$\mathrm{cone}(f)_n := A_{n-1} \oplus B_n \qquad \text{and} \qquad \partial := \begin{pmatrix} -\partial^A & 0 \\ f & \partial^B \end{pmatrix}.$$

Remark 33.15. The literature contains a variety of alternative versions of Definition 33.14 with slightly different sign conventions.

It is straightforward to check that for any chain map $f : A_* \to B_*$, one obtains a short exact sequence of chain complexes

$$0 \longrightarrow B_* \xrightarrow{i} \mathrm{cone}(f)_* \xrightarrow{\pi} A_*[-1] \longrightarrow 0,$$

where we denote by $A_*[-1]$ the chain complex A_* with its grading shifted so that $A_*[-1]_n := A_{n-1}$, and the maps i and π are the obvious inclusion and projection respectively,

$$B_n \xrightarrow{i} A_{n-1} \oplus B_n, \qquad A_{n-1} \oplus B_n \xrightarrow{\pi} A_{n-1}.$$

Plugging this into Proposition 32.13 thus gives a long exact sequence that relates the homology groups of A_*, B_* and the cone, and by inspection of the usual diagram chase, one finds that the connecting homomorphism $H_n(A_*[-1]) = H_{n-1}(A_*) \xrightarrow{\partial_*} H_{n-1}(B_*)$ in this case is simply the map induced on homology by the chain map $f : A_* \to B_*$, so the long exact sequence takes the form

$$\ldots \longrightarrow H_n(A_*) \xrightarrow{f_*} H_n(B_*) \xrightarrow{i_*} H_n(\mathrm{cone}(f)_*) \xrightarrow{\pi_*} H_{n-1}(A_*) \xrightarrow{f_*} H_{n-1}(B_*) \longrightarrow \ldots. \tag{33.4}$$

The exactness of this sequence implies:

Proposition 33.16. *A chain map $f : A_* \to B_*$ induces isomorphisms $H_n(A_*) \to H_n(B_*)$ for all $n \in \mathbb{Z}$ if and only if its mapping cone $\mathrm{cone}(f)_*$ is acyclic.* ☐

The following is a chain-level analogue of Proposition 33.16.

Theorem 33.17. *A chain map $f : A_* \to B_*$ is a chain homotopy equivalence if and only if its mapping cone $\mathrm{cone}(f)_*$ is chain contractible.*

Proof. Suppose $\mathrm{cone}(f)_*$ admits a chain contraction, so for each $n \in \mathbb{Z}$, there is a homomorphism

$$h = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : A_{n-1} \oplus B_n = \mathrm{cone}(f)_n \to \mathrm{cone}(f)_{n+1} = A_n \oplus B_{n+1}$$

satisfying $h\partial + \partial h = \mathbb{1}$, which amounts to the four equations

$$\begin{aligned} -\partial^A \alpha - \alpha \partial^A + \beta f &= \mathbb{1}, \\ f\alpha + \partial^B \gamma - \gamma \partial^A + \delta f &= 0, \\ -\partial^A \beta + \beta \partial^B &= 0, \\ f\beta + \partial^B \delta + \delta \partial^B &= \mathbb{1} \end{aligned}$$

for the maps $\alpha : A_{n-1} \to A_n$, $\beta : B_n \to A_n$, $\gamma : A_{n-1} \to B_{n+1}$ and $\delta : B_n \to B_{n+1}$. The third equation makes β a chain map $B_* \to A_*$, the first makes $-\alpha$ a chain homotopy between $\beta \circ f$ and the identity $A_* \to A_*$, and the fourth makes δ a chain homotopy between $f \circ \beta$ and the identity $B_* \to B_*$, proving that β is a chain homotopy inverse of f.

We will not need the converse in the application below, so we include here only a sketch of its proof, adapted from an argument in [Bro94, Prop. 0.7].

As a preparatory observation, note that for any two chain complexes (A_*, ∂^A) and (B_*, ∂^B), there is a chain complex $(\mathrm{Hom}(A_*, B_*), \partial)$ whose degree n part consists of the homomorphisms

$\varphi : A_* \to B_*$ that satisfy $\varphi(A_k) \subset B_{k+n}$ for all $k \in \mathbb{Z}$, with the boundary operator ∂ given on homomogeneous elements $\varphi \in \mathrm{Hom}(A_*, B_*)$ of degree $|\varphi|$ by

$$\partial\varphi := \partial^B \circ \varphi - (-1)^{|\varphi|}\varphi \circ \partial^A.$$

The choice of sign convention used here is motivated by a convention that we will later use for defining tensor products of chain complexes, and it ensures for instance that the obvious evaluation map

$$\mathrm{Hom}(A_*, B_*) \otimes A_* \to B_* : \varphi \otimes a \mapsto \varphi(a)$$

is a chain map. This detail is unimportant for now; one can easily check in any case that $(\mathrm{Hom}(A_*, B_*), \partial)$ as defined above is a chain complex. Moreover, the 0-cycles in $\mathrm{Hom}(A_*, B_*)$ are precisely the chain maps from A_* to B_*, and two such cycles are homologous if and only if they are chain homotopic.

Next, we have two claims whose proofs are both straightforward exercises:

Claim 1: For any fixed chain complex C_*, there exists a covariant functor $\mathsf{Ch} \to \mathsf{Ch}$ that sends each chain complex A_* to the chain complex $\mathrm{Hom}(C_*, A_*)$ and sends each chain map $f : A_* \to B_*$ to the chain map

$$\mathrm{Hom}(C_*, f) : \mathrm{Hom}(C_*, A_*) \to \mathrm{Hom}(C_*, B_*) : \varphi \mapsto f \circ \varphi,$$

and moreover, the chain homotopy class of $\mathrm{Hom}(C_*, f)$ depends only on the chain homotopy class of f.

Claim 2: For any chain map $f : A_* \to B_*$ and any third chain complex C_*, there is a natural isomorphism between the chain complexes $\mathrm{Hom}(C_*, \mathrm{cone}(f)_*)$ and $\mathrm{cone}(\mathrm{Hom}(C_*, f))_*$.

With these ingredients in place, suppose $f : A_* \to B_*$ is a chain homotopy equivalence, and abbreviate $C_* := \mathrm{cone}(f)_*$. Claim 1 implies that $\mathrm{Hom}(C_*, f) : \mathrm{Hom}(C_*, A_*) \to \mathrm{Hom}(C_*, B_*)$ is then also a chain homotopy equivalence, and by claim 2, its mapping cone is naturally isomorphic to $\mathrm{Hom}(C_*, C_*)$, implying via Proposition 33.16 that $\mathrm{Hom}(C_*, C_*)$ is acyclic. The vanishing of $H_0(\mathrm{Hom}(C_*, C_*))$ means that every chain map $C_* \to C_*$ is chain homotopic to zero: since this applies in particular to the identity map $C_* \to C_*$, it follows that C_* is chain contractible. ☐

Theorem 33.9 in the case $G = \mathbb{Z}$ is an immediate consequence of the following:

COROLLARY 33.18. *For two chain complexes A_*, B_* of free abelian groups, a chain map $f : A_* \to B_*$ is a chain homotopy equivalence if and only if the induced maps $f_* : H_n(A_*) \to H_n(B_*)$ are isomorphisms for all $n \in \mathbb{Z}$.*

PROOF. If $f_* : H_n(A_*) \to H_n(B_*)$ is an isomorphism for every n, then by Proposition 33.16, $\mathrm{cone}(f)_*$ is acyclic. Since the chain groups A_{n-1} and B_n are free abelian groups, the same holds for $\mathrm{cone}(f)_n = A_{n-1} \oplus B_n$, and it then follows via Lemma 33.12 that $\mathrm{cone}(f)_*$ is also chain contractible. The result now follows from Theorem 33.17. ☐

33.6. Exercises.

EXERCISE 33.1. Let Top_* denote the category of pointed spaces with base-point preserving continuous maps, so that we can regard both π_1 and $H_1(\cdot; \mathbb{Z})$ as functors from Top_* to the category Grp of groups with homomorphisms. (Note that the base point is irrelevant for the definition of $H_1(\cdot, \mathbb{Z})$, which actually takes values in the smaller subcategory of *abelian* groups, but these details are unimportant for now.) In this context, show that the Hurewicz homomorphism (33.3) defines a natural transformation from π_1 to $H_1(\cdot; \mathbb{Z})$.

EXERCISE 33.2. A chain complex C_* is said to be **bounded below** or **bounded above** if $C_n = 0$ for all $n \in \mathbb{Z}$ sufficiently small or sufficiently large respectively, e.g. all of the chain complexes that we have used for defining topological invariants so far have been bounded below, since they satisfy $C_n = 0$ for $n < 0$. Show that if C_* is an acyclic chain complex of R-modules that is bounded above or below and the modules $C_n \subset C_*$ are all free, then C_* is chain contractible. *Hint: Construct a chain contraction inductively by degree, as in the method of acyclic models. The assumption that C_* is bounded above or below gives you a place to start the induction.*

Exercise 33.3. Prove that for any two subsets $\mathcal{U}, \mathcal{V} \subset X$ with $X = \mathring{\mathcal{U}} \cup \mathring{\mathcal{V}}$, the obvious inclusion

$$C_*(\mathcal{U}) + C_*(\mathcal{V}) \hookrightarrow C_*(X)$$

is a chain homotopy equivalence.

Bibliography

[Are46] R. Arens, *Topologies for homeomorphism groups*, Amer. J. Math. **68** (1946), 593–610. MR19916

[Art91] M. Artin, *Algebra*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1991.

[Boo77] B. Booss, *Topologie und Analysis*, Springer-Verlag, Berlin-New York, 1977 (German). Einführung in die Atiyah-Singer-Indexformel; Hochschultext.

[BB85] B. Booss and D. D. Bleecker, *Topology and analysis*, Universitext, Springer-Verlag, New York, 1985. The Atiyah-Singer index formula and gauge-theoretic physics; Translated from the German by Bleecker and A. Mader.

[Bro94] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics, vol. 87, Springer-Verlag, New York, 1994. Corrected reprint of the 1982 original.

[Che92] P. R. Chernoff, *A simple proof of Tychonoff's theorem via nets*, Amer. Math. Monthly **99** (1992), no. 10, 932–934.

[DK90] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1990. Oxford Science Publications.

[ES52] S. Eilenberg and N. Steenrod, *Foundations of algebraic topology*, Princeton University Press, Princeton, New Jersey, 1952.

[FW99] G. K. Francis and J. R. Weeks, *Conway's ZIP proof*, Amer. Math. Monthly **106** (1999), no. 5, 393–399.

[Gal87] D. Gale, *The Teaching of Mathematics: The Classification of 1-Manifolds: A Take-Home Exam*, Amer. Math. Monthly **94** (1987), no. 2, 170–175.

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[Hir94] M. W. Hirsch, *Differential topology*, Springer-Verlag, New York, 1994.

[Jän05] K. Jänich, *Topologie*, 8th ed., Springer-Verlag, Berlin, 2005 (German).

[Kel50] J. L. Kelley, *The Tychonoff product theorem implies the axiom of choice*, Fund. Math. **37** (1950), 75–76.

[Kel75] J. L. Kelley, *General topology*, Graduate Texts in Mathematics, vol. 27, Springer-Verlag, New York, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.]

[Lan02] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002.

[Lee11] J. M. Lee, *Introduction to topological manifolds*, 2nd ed., Graduate Texts in Mathematics, vol. 202, Springer, New York, 2011.

[LL01] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics, vol. 14, American Mathematical Society, Providence, RI, 2001.

[Man14] C. Manolescu, *Triangulations of manifolds*, ICCM Not. **2** (2014), no. 2, 21–23.

[MS12] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, 2nd ed., American Mathematical Society Colloquium Publications, vol. 52, American Mathematical Society, Providence, RI, 2012.

[Mil62] J. W. Milnor, *On axiomatic homology theory*, Pacific J. Math. **12** (1962), 337–341.

[Mil97] ———, *Topology from the differentiable viewpoint*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Based on notes by David W. Weaver; Revised reprint of the 1965 original.

[Moi77] E. E. Moise, *Geometric topology in dimensions* 2 *and* 3, Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, Vol. 47.

[Rad25] T. Radó, *Über den Begriff der Riemannschen Fläche*, Acta Szeged **2** (1925), 101–121.

[Rud87] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1987.

[Spa95] E. H. Spanier, *Algebraic topology*, Springer-Verlag, New York, 1995. Corrected reprint of the 1966 original.

[Ste67] N. E. Steenrod, *A convenient category of topological spaces*, Michigan Math. J. **14** (1967), 133–152.

[Tho54] R. Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86 (French).

[Tho92] C. Thomassen, *The Jordan-Schönflies theorem and the classification of surfaces*, Amer. Math. Monthly **99** (1992), no. 2, 116–130.

[Wen18] C. Wendl, *Holomorphic curves in low dimensions: from symplectic ruled surfaces to planar contact manifolds*, Lecture Notes in Mathematics, vol. 2216, Springer-Verlag, 2018.